DOCUMENT RESUME

ED 263 695 EC 180 962

AUTHOR	Anderegg, David; And Others
TITLE	Strategy Training Project. Final Report.
INSTITUTION	Research Inst. for Educational Problems, Cambridge, Mass.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	Jun 85
GRANT	G00820060
NOTE	130p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Classification; *Cognitive Processes; Elementary Education; Reading Comprehension; *Reading Difficulties; *Reading Instruction; *Transfer of Training

ABSTRACT

Forty-seven reading disabled children whose major reading problem was comprehension participated in a study to examine the effects of interventions designed to structure information in encoding and recall through the use of categorization strategy. Short term treatment consisted of four tutorials over a 2-week period. Long term treatment 1 consisted of eight tutorials over 4 weeks (an extension of the short term treatment). Long term treatment 2, also eight tutorials over a 4-week period, included material from content areas studied in the regular classroom. A control group was not exposed to the strategy training curriculum or any other type of group experience. Measures of clustering and recall were used to assess the impact of the strategy training curriculum. There were no differences in the degree of categorization evident in the recall protocols of trained and control Ss with respect to picture stimuli. There were also no significant differences between Ss in long term and short term training conditions regarding adjusted ratio of categorization (ARC) scores. Comparisons between treatment and controls showed no signficant effects of training, and extent of training had no significant effect on recall of items from either blocked or unblocked prose stimuli. Practical and design-related concerns were noted to affect data collection and research results. It is noted that "uninsightful" poor readers (i.e., those requiring additional steps in the training process) did improve as a result of their exposure to strategy training. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED263695

FINAL REPORT

STRATEGY TRAINING PROJECT

Grant G00820060

David Anderegg, Mark Meterko and Milton Budoff
Research Institute for Educational Problems
Cambridge, Massachusetts

Submitted to
U.S. Department of Education, Office of Special Education
June 1985

The research reported herein was performed pursuant to a grant from the U.S. Department of Education, Office of Special Education. Grantees undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official U.S. Department of Education policy.

EC180962

BEST COPY AVAILABLE

TABLE OF CONTENTS

Page

I. INTRODUCTION 1

II. METHODS 17

A. Overview 17

B. Subject Selection 18

C. Design 27

D. Measures 28

1. Overview 28
2. Clustering in Free Recall 29
3. Recall Efficiency 33

E. Hypotheses 34

III. RESULTS 38

A. Overview 38

B. Categorization of Pictures 40

1. Control vs. All Treatment Groups 45
2. Control vs. Short Term 48
3. Short Term vs. Long Term 49

C. Categorization of Prose 49

1. Control vs. All Treatment Groups 52
2. Control vs. Short Term 52
3. Short Term vs. Long Term 55

D. Relationship Between Categorization and Recall Efficiency 56

1. Recall Efficiency 56
    a. Recall of Pictures 56
    b. Recall of Prose 60
2. Categorization and Recall 64

E. Insightful and Unsightful Children 67

1. Definition and Distribution 67

F. Note Regarding Gender Differences and Digit-Span Capacity Effects 73

1. Overview 73
2. Sex Effects 73
3. Digit-span Capacity Effects 74

TABLE OF CONTENTS

(continued)

Page

IV. DISCUSSION 79

A. Practical Issues Bearing on Study Outcome 82

1. School Climate 82
2. Peer Contamination Effects 83

B. Design-Related Issues Bearing on Study Outline 84

1. Category Cueing and the Effects Pre-test Procedures ... 84
2. Memory Load 87
3. Group and Individual Differences 89

REFERENCES 93

APPENDICES:

A. Letter to Teachers, Including Description of Poor Readers 96

B. Strategy Training Curriculum Manual: Pilot Study Version 98

INTRODUCTION

Interest in information-processing strategies as determinants of intelligent performance has increased dramatically in the last decade. A wide variety of theoretical issues and attendant empirical studies have been undertaken, based upon a conceptualization of the human information-processing system that distinguishes between the routines used for processing and storing information on one hand and the "contents" of the information-processing system on the other. In simple terms, it is currently widely accepted that it is more important to "know how to know" than simply to "know." Adaptive approaches to one's environment can be facilitated by enhancing routines for problem-solving, remembering, etc., rather than by filling slots in the information-processing system without providing the routines or strategies necessary to access information already stored and to efficiently store incoming information.

In addressing issues of information-processing strategy use by human knowers, the central question for developmentalists and educators concerns the acquisition of information-processing strategies. How are these strategies acquired, and at what age? What are the constraints imposed upon the ideal information-processing system by immaturity or handicapping conditions? How can these constraints be conceptualized in studying, and eventually ameliorating, the information-processing deficits of immature knowers?

A useful distinction has been made between control processes and structural differences in the performances of human knowers. Initially drawn by Atkinson and Schiffrin (1968), this distinction has come to be widely accepted in the field of cognitive psychology. In essence, "control processes" refer to optional strategies that an individual brings to bear upon a task (a memory task, a problem-solving task, etc.). Structural differences refer to the invariant components of the system that constrain the ability of the problem solver to perform a given task. To make an analogy with computer operation, the structural differences are likened to computer hardware--for example, the number of "bits" that can be stored in the system. Control processes might be likened to software, in that they operate upon information to make for efficient problem solving. The efficiency of the system is greatly affected by the software (for example, the number of steps in a program), but the capacity of the system is ultimately constrained by some hardware features.

A major implication of the hardware/software distinction is that control processes are susceptible to training, while structural features are not (Campione & Brown, 1977). This distinction is essentially the same as that made by Flavell (1970) in his discussion of strategies in memory tasks in which he compares "production deficiencies," or the lack of spontaneous production of mnemonic strategies, and "mediation deficiencies," or the lack of ability to use mnemonic strategies when provided. Volumes of research data have been generated during the past ten years as a result of Flavell's critique by studies attempting to determine which tasks can be learned (and thus are subject to control processes) and which tasks cannot be learned (and thus are constrained by structural features), by subjects of varying ages (Flavell and Wellman, 1977) as well as by subjects with handicapping conditions, notably mental retardation (Fisher and Zeaman, 1973; Campione and Brown, 1977).

Another major implication of the control processes-structural limitations distinction is the necessity for considering developmental models of cognitive processes. Structural limitations may be temporary (Campione and Brown, 1977), and what may be trained at one developmental level is not the same as what may be trained at another developmental level. Each of the structural components described by a particular information processing model may be a locus of change that affects the trainability of control processes at different stages of individual development. For example, one model of intellectual processes (Brown and Campione, 1979) introduces the distinction between features of short-term memory--including capacity (the number of "slots" in short-term memory), durability (the length of time of decay of information in short-term memory) and efficiency (various other processes, like speed of access, etc. in short term memory)--and features of long-term memory, like knowledge base (the actual quantity of experience of the knower), schemes (Piagetian-type rules of thinking) and control processes (rules and strategies for knowing, remembering, problem-solving, etc.).

These general areas may, of course, be described using different terms, depending upon the sub-specialty of the author. The degree of emphasis on one or another component of the information-processing system also differs from one account to another. Neo-Piagetians (Pascual-Leone, 1970; Case, 1978), for example, place a great deal of emphasis upon working capacity and its relation to operative schemes, and correspondingly less emphasis upon control processes or metacognitive components of the information-processing system. The critical point is that all these components may develop; capacity, durability, efficiency, knowledge base, schemata and control processes can, theoretically, change in the course of ontogenesis and can serve as structural limitations on other parts of the developing information-processing system. For example, Chi (1978) discusses ways in which limitations of the knowledge base can account for differential performance by children and adults on memory tasks in experimental situations in which mnemonic strategies have been prevented (that is, neither group is using mnemonic strategies). In such a situation, the size of the knowledge base is a structural limitation upon the performance of the memory task. Both groups, if trained to use strategies or allowed to spontaneously use strategies, would be expected to improve their task performance, but the relative position of children and adults would remain constant. Similar accounts have been part and parcel of Piagetian literature for some time: differences in operational level provide structural limitations on the performance of children that are resistant to training efforts (Kuhn, 1974).

In sum, the operational definition of control processes as those strategies or rules for problem solving that can be trained is subject to two major limitations: (1) failure to perform adequately on any particular training assessment may be the result of a faulty training procedure; and (2) failure on a particular training assessment may be the result of a temporary structural limitation due to the immaturity of the information-processing system. In spite of these limitations, the distinction between control processes and structural limitations is valid because of its power to distinguish between what can be changed and what cannot. The emphasis becomes not "what a person can do" but "what a person can be *trained* to do," a distinction with critical importance when discussing the problem of poor learners.

The training paradigm has recently been discussed in theoretical terms as the "instructional approach" to cognitive psychological research. Belmont and Butterfield (1977) set forth the criteria for the ideal training study in psychological research, including provisions for direct measurement of strategy use, detailed task analysis of the problem to which the strategies are to be applied, and well-defined evaluation criteria to assess strategy-use competence. In addition, they note that failure to see results in the implementation of a training study may be the result of inadequate training procedures. Indeed, the requirements of the instructional approach are such that definitive statements about the nature of a production or mediation deficiency are not possible without *repeated* attempts to train the strategy in question. As Turnure, Buium and Thurlow (1976) point out, some inefficient information-processing performances may be the result of an "instructional deficiency": the experimenter has not been sufficiently ingenious in eliciting spontaneous strategy production or in training strategies for information processing. This has the effect of "blaming the victim" for a less-than-adequate performance in an experimental testing when, in fact, the experiment itself does not allow for the strategy to be implemented by the subject

(Borkowski and Cavanaugh, 1979).

An additional criticism of training studies in general has been their superficial quality. As Brown and DeLoache (1978) point out, literally hundreds of training studies in information-processing strategies have appeared in recent years, but their quality with respect to actually improving the performance of the subjects has been less than inspiring. The instructional relevance of the materials to be processed (remembered, etc.) is negligible; the strategies have been trained, but the effect on the subject is not long-lasting. As these reviewers point out, repeated demonstration studies will not add anything further to our knowledge about the production of strategies by inefficient knowers or the amenability to training of these knowers. We know that inefficient information processors do not spontaneously use strategies, and we know they can be trained to do so. What is been needed is research on the applications of these findings.

Since the early discussions of the production-mediation deficiency hypothesis, these constructs have figured prominently in the discussion of the information-processing deficits of the handicapped. Ellis (1970) discussed the differences between retarded and normal populations as being due to a difference in the spontaneous use of information-processing strategies. Because of the lack of improvement in retarded persons' performance on a serial recall task as a function of increased exposure time, Ellis hypothesized that normal subjects were doing something with the increased time that retarded subjects were failing to do; specifically, normal subjects were rehearsing the stimuli to be recalled and the retarded subjects were not.

Subsequent research with the retarded lends support to this interpretation. In two early studies, retarded children were trained to rehearse stimuli to be remembered on a serial recall task. Early rehearsal training studies with EMR populations demonstrated that this strategy can, in fact, be trained and the strategy can be maintained over time by this population. In one such training study, Brown, Campione, Bray and Wilcox (1973) trained EMR adolescents with a mean MA of 7.5 years to rehearse an inspection set of stimuli in a "keeping-track task" (a task in which the subject is required to monitor the current state of a number of variables as they are changed). Retardates trained to use a cumulative rehearsal strategy--that is, saying to themselves the stimuli in a set of cumulative fashion ("cat; cat-dog; cat-dog-bear; cat-dog-bear-lion," etc.)--performed the keeping track task about as well as normal adolescents. In a similar experiment, when normal subjects were prevented from rehearsing by contracting the inter-trial time interval, they performed about as well as untrained retarded subjects. These results were replicated six months after training without prompting by the experimenters (Brown, Campione and Murphy, 1974), although it must be noted that the setting, experimenter, and stimuli were all retained to provide for maximum familiarity for the retarded subjects.

Attempts to train a more complex rehearsal strategy to non-producing subjects were made by Butterfield, Wambold and Belmont (1974). These attempts were generally successful, in that the retarded subjects (CA ranging from 13-21) did learn to use the trained strategies. However, deficienies in sequencing and selecting sub-routines required the training to be highly task-specific. For example, subjects were trained to use a "cumulative rehearsal-fast finish" strategy to remember the positions of letters in a six-letter list. The strategy involves (1) rehearsing and storing the first three letters, (2) keeping the last three letters in short-term, or passive, memory, and (3) attempting to answer the position probe before the memory trace erodes from short-term storage. In the course of training, the subjects were taught to perform eight separate steps:

1. Use the passive system to construct a rehearsable group of the first three letters
2. Transfer these letters to the active memory store by rehearsing them
3. Stop attending to those first three letters while exposing and attending to but not rehearsing the last three letters
4. Expose the probe letter immediately after attending to the sixth letter
5. Immediately search the entire contents of the passive memory store
6. Respond at once if the probe is found in the passive store
7. If not found, search the active store serially
8. Terminate that search by responding when the probe item is located.

Performance on the memory task did not improve when subjects were trained to perform steps 1, 2 and 3 together, or when they were trained to use steps 1, 2, 5, 6, 7 and 8 together. Only when they were trained on all eight steps in sequence did the retarded subjects' memory performance resemble the performance of normal subjects.

The complexity of the instruction needed to produce gains in this study seriously raised the question of the generalizability of strategy training. If such complex training is needed for such a simple task, can the training ideal ever proceed to "real life" tasks, which are infinitely more complex? More recently, researchers in the area of strategy training have responded to this question by attempting to train strategy generalization using a rehearsal strategy. As noted above, one approach to the failure of a training study is to alter the training procedure: if the subject fails to perform adequately after training, the training procedure itself may be at fault, not the subject's ability to learn. Belmont and Borkowski (1978) demonstrated that retarded children can be taught to use a rehearsal strategy that generalizes to highly similar, but not identical, recall tasks. In learning 7-letter strings, various input strategies were trained in combination: cumulative rehearsal of the first four, followed by passive storage of the last three; cumulative rehearsal of the first three, followed by passive storage of the last four, etc. In tasks requiring the recall of the string in various orderings (last three, then first four; last two, then first five, etc.), retarded children who had received two training sessions performed better than once-trained children, especially in later trials. It appears that the amount of training and the teaching of multiple strategies facilitated performance on an untrained, but very similar, recall task.

The utility of training information-processing strategies to retarded subjects is not limited to tasks of serial recall. The vast training literature cannot be reviewed in its entirety here. Efforts to train handicapped subjects to use strategies of categorization and clustering as aids to recall have been reviewed by Bilsky and Evans (1979); and the general area of strategy training with the handicapped learner has been reviewed by Brown and Campione (1977). The bulk of these studies suggest that training information-processing strategies can result in greatly improved performance by the handicapped learner on a wide variety of information-processing tasks.

In research on strategy maintenance and generalization by the retarded, the amount of training involved has been found to be a critical factor. The question of maintenance and generalization has proven to be one of the most difficult to address with retarded populations because of definitional issues about what types of performances constitute evidence for maintenance--that the strategy has been retained in substantially unchanged form after the training sessions--and generalization--that the strategy has been retained and has been applied to new situations by the trained subject. Current workers in the field of strategy training disagree about the extent of both maintenance and generalization of strategies trained to retarded subjects. Some reviewers, notably Campione and Brown (1977) conclude that maintenance of strategies has been demonstrated, especially in cases where a sufficient amount of training was involved. Other reviewers, notably Borkowski and Cavanaugh (1979) find evidence for both maintenance and generalization of trained strategies by retarded subjects.

Two issues must be raised with respect to the question of the possibility of long-lasting effects of strategy training with retarded subjects. The first issue, mentioned above, is that the failure to demonstrate generalization effects should by no means be regarded as testimony to the subjects' limitations until a wide variety of training procedures have been employed. In other words, the "instructional deficiency" of experimenters described by Turnure, Buium, and Thurlow (1976) may account for the lack of impressive generalization effects by retarded subjects trained in information-processing strategies reported thus far. The second issue, raised by Borkowski and Cavanaugh (1979) in their review, is that generalization is not likely to be demonstrated in training situations that do not maximize maintenance. Strategy maintenance is a logical prerequisite for strategy generalization. Training methodologies that attempt to demonstrate generalization must therefore maximize the probability of strategy maintenance. One way to maximize strategy maintenance is to increase the amount of strategy training.

This proposal has been supported by a number of recent studies that have involved manipulation of the amount of training sessions to assess the impact of amount of training on strategy maintenance. Number of sessions has been found to significantly increase strategy maintenance with elaboration strategies on a paired-associate task by retarded children (Turnure and Thurlow, 1973). Borkowski, Cavanaugh and Reichart (1978) also found a significant increase in strategy maintenance due to increased number of training sessions with a sample of normal children. In both these cases, the significant gains in strategy maintenance were made after an increase from one to two training sessions. Although individual differences in maintenance also occur that are related to the "level" of training, it appears that at least one necessary condition for the demonstration of strategy maintenance is sufficient amount of training for the subject to acquire the strategy in a meaningful way. We therefore conclude that long-term strategy training is a necessary precondition for the maintenance and generalization of information-processing strategies.

The research reported here responds to the critical issues raised in the previous section. Although the training study as a research paradigm has generated a great deal of sophisticated training technology in the area of cognitive psychology, little of this sophistication has been passed on to workers in the field of education. In part, this lack of communication may be attributed to the increasing specialization in both psychology and education. Little is known on either side about the others' work, and this isolation often results in a duplication of efforts. As Brown (1978) reports in her review of the relationship between metacognition and information-processing strategies, the entire area of metacognition and strategic approaches to information-processing was at one time known as "study skills" (Robinson, 1941).

In general, research in cognitive psychology has been rigorously controlled with respect to appropriate experimental procedure, including the careful manipulation of relevant independent variables, appropriate control populations, and so on. On the other hand, research in the application of strategy training has been characterized by instructional relevance but a lack of attention to appropriate experimental design and the detailed specification of training procedures and manipulations of relevant variables in the training domain (Borkowski and Cavanaugh, 1979). The objectives of the research reported here were, in general, to address this knowledge gap with some experimental research that is instructionally relevant: to attempt to move from the laboratory to the classroom without losing the lessons learned in the laboratory. This goal, while difficult to achieve in practice because of the specialization of the professionals involved, nevertheless remains the logical extension of the efforts at strategy training. As stated above, we now know that retarded subjects can be trained to use information-processing strategies, but can they be integrated into their daily knowledge routines? In other words, the retarded can be trained to use strategies, but can they be trained to find them useful? In their seminal review of strategy training literature for the retarded, Borkowski and Cavanaugh state the problem as follows:

> It is clear that the field of instructional research needs well-controlled outcome studies aimed at assessing the advantages and disadvantages of incorporating strategy-based training programs into formal educational curricula. Since applied instructional studies have not been systematically conducted with proper attention to controls, we are working more from faith than fact when we urge teachers to tell their children how to learn as well as what to learn. The implications of this maxim for educational change are so great that the facts must be gathered in the decade ahead. [p. 588]

Consequently, the objectives of the research reported here were twofold:

1. To increase the dialogue between the disciplines of cognitive psychology and education by applying training technology from cognitive psychological training studies to educationally relevant problems in the classroom; and

2. To test the hypothesis that long-term support for the use of information-processing strategies will result in more maintenance and generalization of trained strategies than will short-term interventions.

In order to meet the first objective, a cognitive strategy was needed to be selected that was (a) well-documented in the cognitive psychology training literature, and (b) was capable of being adapted, with little modifications, to tasks that might be encountered in a school setting. Categorization was selected because of its extensive use as a model of an information processing strategy in cognitive research, and because of its ready application to a familiar classroom task: remembering items from passages of prose. Although many categorization training studies have involved the memorization of stimuli from lists of arrays of pictures, a simple translation of these items into a prose format makes the task much more similar to a task that students encounter every day in the classroom. Remembering items from a prose passage can be greatly enhanced using the strategy of categorization. Such a task is not, however, identical to a categorization of items in list form. For example, the step of "disembedding" items from the prose passage, or selecting items to be remembered and ignoring the connecting prose, is the feature of the task that makes it dissimilar enough from list learning to be a "transfer task" that will test the ecological validity of the strategy of categorization.

It should be noted here that a burgeoning literature has arisen in the field of cognitive psychology concerned with the information-processing requirements of reading comprehension vis a vis the "instructional approach"; that is, which uses training studies to illuminate the basic psychological processes in reading comprehension (Brown, Campione & Day, 1981). While we agree with this approach and find it compatible with our selection of prose passages as a transfer task for categorization, it should be noted that we do not claim to encompass the entire complexity of the phenomenon of reading comprehension in this study. Because prose passages were selected as the transfer task of interest, we used a population of poor readers as our target population, including many children who were labeled "learning disabled" or "reading disabled" by the schools they attended. The use of the strategy of categorization for items embedded in prose passages was regarded primarily as a test of the transfer of the basic strategy. Although we do believe that cognitive strategies, including categorization, have much to do with the reading comprehension process, this process is complex; it was not our expectation that all the problems of the poor readers selected for this study would be ameliorated by the curriculum designed and evaluated here.

To meet the second objective of the present research, a strategy training curriculum was designed that included a short-term introduction to the idea of categorization and its application to passages of prose. A long-term curriculum was also designed that consisted of twice the number of tutorial sessions and reinforced the lessons in the short-term curriculum. The long-term curriculum applied the categorization strategy to longer and more complex prose passages in an effort to test the limits of the subjects' ability to maintain and generalize the trained strategy. Two versions of the long-term curriculum were designed, to differ in one respect: in one condition, reading materials were chosen from the subjects' regular school classrooms (that is, assignments were taken directly from their social studies or science books). This condition was added so as to explicitly test the conditions necessary for transfer to school-relevant tasks.

The research reported here, then, addressed the failure, described in the literature of psychology and education, to apply the sophisticated training methods and concepts from cognitive psychology to real problems in education of poor learners. The strategy of categorization, the transfer task of remembering items embedded in prose, and the population of poor readers were selected as a coherent whole to achieve this end. It was felt that these represented a well-documented cognitive strategy with a history of successful training applications; a transfer task that was likely to be encountered by handicapped learners in an actual school setting; and, finally, a population in need of appropriate interventions.

METHODS

Overview

The strategy training project attempted to improve the comprehension skills of poor readers in the sixth and seventh grade. Three major issues were addressed: (a) could these children be taught to use categorization strategies for remembering information? (b) could they then learn to transfer this skill from the original training materials (stacks of pictures) to prose paragraphs? and (c) if they could learn to apply the cognitive strategies to prose, did those strategies in fact increase their retention and comprehension of what they read? Children nominated by their teachers were randomly assigned to one of three treatment groups or a control group.

In each treatment condition the subjects met individually with a strategy training tutor. Each meeting required 30 to 90 minutes, depending on the child. Training sessions averaged between 40 minutes and one hour in length. The training curriculum is described in detail elsewhere in this report. The treatment groups differed with regard to (a) the overall duration of the project, and (b) the types of materials used. One group met regularly for 4 weeks; the other two treatment groups met for 8 weeks. Two groups used practice materials that had been designed specifically for the strategy training curriculum; 1 treatment group incorporated readings based on material that the subjects were currently studying in their regular classrooms.

To assess the effectiveness of the strategy training curriculum and the different treatment approaches, the subjects were tested at two points in time: (1) a few days prior to the first regular meeting with the tutor, and (2) within a week following the last tutoring session. Children in the two long term treatments were also tested at the midpoint of their training cycle at a time corresponding to the end of the short term treatment. In summary, we used a pretest-posttest control group experimental design. The control group was a no-treatment condition. The details of subject selection, experimental design, and the dependent measures used to assess the acquisition and use of the categorization strategies are described below.

Subject Selection

Children who participated in the strategy training project were recruited from four classrooms--two sixth grade and two seventh grade--at the host school. The school was a public school in a working-class suburb of Boston, with a population ranging from poor, not working families up to middle-class families, with some white collar occupations represented in the upper levels of income and status. This particular school was located near a large public housing project, and students from the project made up a substantial minority of the school's population. The subject population, then, was drawn from a larger population in which optimal conditions for good education do not prevail; indeed, some of the paramount concerns of the school staff center around attendance and truancy, which is a significant problem at such an institution.

This institution, then, contributed to the types of learning problems of some of the experimental subjects (see below) and contributed as well to some difficulties in data handling (for example, one of the experimental subjects, when he discovered that he was not being promoted to the eighth grade, left school for the remainder of the year, making post-test data collection impossible).

Since the project was aimed at remediating the problems of "poor readers," criteria were adopted to select a subject population who were deficient in the skills to be trained in the experimental treatment, that is, remembering items listed in a prose format. Two criteria were adopted: (1) that the children who were selected were "poor readers" as defined by reading at least two grades below their actual grade level; (2) that these children did not have any significant problems in decoding, that is, they were able to read aloud and extract meaning from prose passages.

The initial proposal for this research focused on EMR children, because much of the initial work in cognitive strategy training was done using EMR subjects. The subject population for this research changed during the course of implementation for a number of reasons. First, it was very difficult to find a large sample of EMR children in one place. Because of widespread mainstreaming and non-labelling policies in Massachusetts, there are few labelled EMR children remaining in substantially separate classrooms in schools in the Boston area. To achieve a large enough subject population, a great deal of effort would have had to be made to recruit subjects from a number of schools, with concomitant expense in staff time, travel time, and so on. The subjects thus recruited would have been from a large number of schools, which would have increased the problem of subject compatibility, in terms of SES, school climate, etc. It was also the case that, as noted above, the fit between the strategy of categorization and the transfer task of remembering items from prose passages seemed to be the best test of a real-world application of cognitive strategy training. To use this strategy and transfer task, other reading problems--especially decoding problems--had to be ruled out as a source of subjects' poor memory performance. Locating subjects in substantially separate classes who could read without decoding problems proved to be next to impossible. It was assumed that a population in a low SES school, using poor readers in regular education classrooms, might yield a large enough sample of subjects within the walls of one school who might manifest cognitive strategy deficits leading to problems in reading comprehension. In sum, this population was comprised of students labelled "reading disabled" and "learning disabled"; there were, on the basis of PPVT IQ scores, a few subjects who meet the IQ criteria for EMR status.

The selection criteria adopted had some implications for the subject population at large which may have had impact upon the response of subjects to experimental treatment, as follows:

Criterion 1: The selection of "two grades below level" does not select children whose sole learning problem is reading. All the children were in regular education classrooms, but a large majority of the children in the study were receiving special education services for at least their reading problem. A breakdown of subjects by treatment group with respect to their special education status is reported in Table 1.

The adoption of Criterion 1 yielded a subject population that cut across more standard labels used to describe populations of poor learners; the best description of the population selected here is "reading disabled." Formal full-scale measures of intelligence were not available for these subjects since they were not in school records and the time required for individual administration of these measures was prohibitive. However, a rough estimate of overall IQ was generated by giving each subject the Peabody Picture Vocabulary Test. (Although the use of the PPVT as an IQ equivalent is far from ideal, there is substantial precedent for this equation in recent research literature in cognitive psychology.) The range of PVIQ for the subjects in the current study was 64-133. There were, then, subjects who may fall into the traditional "EMR category" by virtue of a low overall IQ and a reading problem. Most of the subjects in the current study fit current and traditional definitional criteria for "learning disabled" in that, with an IQ within normal range, they have specific deficits in one of more areas of school functioning, being two or more grades retarded in reading. By describing these subjects as "reading disabled," we attempt to describe the specific focus of their deficit in school function, and make no assumptions about the etiology of this disability. While many descriptions of LD children rely upon notions of organic impairment and/or developmental lag (Worden, 1983), it is possible that the lag in reading ability observed in our subjects was due to environmental handicapping conditions. Many children in this study came from chronically disorganized famililes without economic resources, with concomitant problems of lack of structure, expressed concretely in the environment in problems in school attendance, illnesses, frequent moves, etc., and expressed by the subjects in their disorganized approach to school tasks and materials. This categorizattion also meets the Federal definition of learning disabled for P.L. 94-142.

Criterion 2: The selection of children whose major reading problem is not decoding but comprehension per se suggests that these children can respond to an experimental intervention aimed at structuring information in encoding and recall. As has been noted recently in descriptions about learning to read, there occurs a shift about the fourth grade when children are no longer asked to read to confirm what is already known--the "learning to read" stage--but are asked to use the tools of written language to take in new information; they are asked to "read to learn" (Chall, 1979). Children who are still "learning to read" were inappropriate for the present intervention in that the focus is on training memory strategies in which the input stage (i.e., getting the information in understandable form) is already routine. The use of this second criterion may, in fact, skew the population in the study away from those with organically based learning disabilities (i.e., dyslexia) and toward a more "environmentally handicapped" population (see above).

A description of the two criteria to be used in subject selection was distributed to teachers (see Appendix 1). At the time of subject selection, standardized reading scores were not available for these subjects, so the teachers' judgment of how the students met the selection criteria was the only source of information available. After the completion of the study, standardized reading scores were obtained, and the results of this testing on the Metropolitan Achievement Test (given 4/15/83) substantially corroborated the teachers' judgment of their pupils' reading abilities (see Table 1).

Using these criteria, 47 subjects were selected for inclusion in the current study. The children were divided into three experimental groups and one control group on the basis of their PVIQ score, with groups matched on this variable. There were no significant differences on PVIQ among the treatment groups (see Table 2). Groups were also matched for sex of child, with boys predominating in the total sample, and thus in each treatment group. Treatment conditions, consisting of one short-term treatment, two different long-term treatments, and one control group, were randomly assigned to the matched groups of subjects (see Table 3).

TABLE 1

Reading Status of Study Participants by Treatment Condition

	Control	Short Term	Long Term 1	Long Term 2	All Subjects
Mean Reading Achievement Grade Level (N)	4.4 (7)	4.3 (11)	4.6 (11)	4.3 (12)	4.4 (41)
n of children not two grade levels behind	1	2	3	2	8 (17%)
n of children receiving additional sp. ed. services	5	5	6	6	22 (47%)

Note: Reading achievement grade level according to metropolitan achievement tests/reading, administered in the host school in April 1983. Study occurred in May, June 1983; all students were in 6th and 7th grade regular classes.

TABLE 2

Treatment Group and Gender Differences in IQ at Time 1: Analysis of Variance Results

		Treatment Group				
		Control	Short Term	Long Term 1	Long Term 2	
Gender	Male	93.8	92.3	94.7	94.9	94.0
	Female	84.3	86.4	87.5	95.7	89.4
		91.2	89.6	92.1	95.2	92.2

Treatment Group F (3,38) = 0.24

Gender F (1,38) = 0.92

Treatment by Gender F (3,38) = 0.20

Table 3

Distribution of Subjects
Within Treatment Conditions
by Grade and Gender

Experimental Group	Grade and Sex				
	Sixth		Seventh		
	Males	Females	Males	Females	
Group 1 Short term	1	2	5	3	11
Group 2 Long term, special materials	3	4	4	0	11
Group 3 Long term, classroom-based	1	2	6	4	13
Group 4 Control	5	1	4	2	12
Totals	10	9	19	9	47

An additional teste was administered at pre-test to provide information regarding baseline short-term memory abilities: the Digit Span subtest of the WISC-R. However, this variable was not considered when matching subjects to create treatment groups, and a significant difference was discovered after the treatments began--the control group was significantly higher on WISC-R Digit Span than the other three (experimental) groups (see Table 4). Subsequent analyses demonstrated that this discrepancy did not affect the subjects' response to experimental treatment.

Table 4

Treatment Group and Gender Differences
in WISC-R Digit Span at Time 1:
Analysis of Variance Results

Treatment Group	Gender		Group Means
	Male	Female	
Control	5.6 (9)	4.0 (3)	5.17 (12)
Short Term	4.3 (6)	4.4 (5)	4.36 (11)
Long Term 1	5.1 (7)	4.0 (4)	4.72 (11)
Long Term 2	4.3 (7)	4.2 (6)	4.23 (13)
Gender Means	4.89 (29)	4.16 (18)	4.62 (47)

Treatment Group	F(3,39) = 2.09+
Gender	F(1,39) = 3.98*
Group by Gender	F(3,39) = 1.55

* p<.05

+ Post hoc comparison of group means (duncan multiple range test, alpha=.05) indicates greater than chance difference between control group and long term 2 treatment group.

Design

A pretest-posttest control group experimental design (Campbell & Stanley, 1963) was used to test the hypotheses of this study. Four groups with roughly equivalent average PPVT scores were created. We then randomly assigned each of these groups to one of the four treatment conditions. The treatment conditions varied with regard to (a) the number of meetings each child had with his/her tutor, and (b) whether classroom readings were integrated into the strategy training curriculum. Specifically:

Group 1: _The short term treatment_ consisted of four tutorial sessions over a two week period. Prose materials used were those prepared specifically for the strategy training project.

Group 2: _Long term treatment 1_ consisted of eight tutorial sessions over a four week period. Prose materials used were prepared specifically for the strategy training project. This treatment thus represented an extended version of Group 1.

Group 3: _Long term treatment 2_ consisted of eight tutorial sessions over a four week period. Prose selections used in this group included material that the subjects were currently studying in their regular classrooms.

Group 4: _Control group_. The group of subjects assigned to the control condition were not exposed to the strategy training curriculum or any other type of group experience. Thus, a no-treatment control rather than a placebo control was used. At the conclusion of the project, the children in the control group did meet once with a tutor, at which time the categorization strategy was explained to them.

By using these four groups we were able to generate the data needed to answer the key questions of this study. A comparison of the control group with the treatment groups would determine whether the strategy training curriculum had any effect on the retention and comprehension of these poor readers. By comparing the performance of the short-term with the long-term treatment groups, we would determine how much exposure was necessary to effectively teach the categorization strategies. Finally, by comparing the long-term group that used only the special curriculum materials with the long-term group that used classroom-based materials, we hoped to discover how much assistance was necessary to insure transfer of learning from the special training experience to the actual classroom environment in which the new skills are to applied.

The specific measures used to assess the students' grasp and application of the categorization and memorization strategies are described next.

Measures

Overview

Two general types of measures were used to assess the impact of the strategy training curriculum: a measure of clustering in free recall, and a measure of recall (i.e., amount remembered). The clustering measure was applied to the category sequence of the items that the child remembered both from the piles of pictures and the prose paragraphs. The amount of information retained by the subjects was calculated as the percent of items recalled for the picture tasks and the prose tasks. These measures are described in detail below. The use of the measures to test the hypotheses regarding the effectiveness of the strategy training curriculum is described in the RESULTS chapter of this report.

Clustering in Free Recall

A key indicator of the effect of the strategy training curriculum was the extent to which the subject grouped items according to category membership. We regarded the child's tendency to cluster items in a free recall task as indicative of that child's memorization (storage) strategy.

Measures of clustering in recall, pioneered by Bousfield (1953), have undergone substantial methodological improvement over the years. We used the adjusted ratio of clustering (ARC) described by Roenker, Thompson and Brown (1971). This measure has several substantial advantages over other clustering mesures, such as the modified ratio of repetition (MRR; Bower, Lesgold & Tieman, 1969), the clustering (C) index (Dalrymple-Alford, 1970), and the deviation (D) index (Dalrymple-Alford, 1970). For example, there is no fixed upper bound for the derivation index; perfect clustering could be represented by different scores, depending on the specific recall protocol. The chance level of category repetition is not set at zero in the computation of MRR and C scores. Therefore the _same_ score may indicate clustering which is above chance for one subject, at a chance level for a second, and below chance for a third.

The ARC score represents the proportion of actual category repetitions above chance to the total possible category repetitions above chance for any given protocol. The ARC score is invariant with respect to factors unrelated to amount of clustering such as the distribution (equal or unequal) of items recalled across the categories represented in the recalled material. A chance level of category clustering is indicated by an ARC score of zero, perfect clustering by an ARC score of 1.0.

The computational formula for the ARC score is:

$$\mathrm{ARC} = \frac{R - E(R)}{\max R - E(R)}$$

where

R = total number of observed category repetitions (i.e., the number of times recalled items followed an item from the same category);

maxR = maximum possible number of category repetitions, given the number of items recalled by the subject and the number of categories represented in the recalled material; maxR is calculated as follows:

$$\max R = N - k$$

where N = total number of items recalled, and
k = number of categories represented; and

E(R) = expected (chance) number of category repetitions, calculated as:

$$E(R) = \frac{\sum_i n(i)^2}{N} - 1$$

where n(i) = number of items recalled from Category i, and, as before, N = total number of items recalled.

Table 5 illustrates the computation of the ARC score for three subjects from the present study. These subjects represent low, moderate, and high degrees of categorization of free recall material. For this study, ARC scores were calculated on the basis of each subject's recall of both picture and prose material.

Sometimes items that were not included in the deck of pictures or not mentioned in the test paragraphs were "recalled" by the subject. These "unofficial" items were regarded as "not of the same category" for the purposes of calculating the number of category repetitions in a recall protocol. As described below, however, unofficial items were counted in the calculation of the memorization score if the subject had included that item in his/her study list for that task.

Table 5

Examples of the Calculation of ARC Scores for Subjects Exhibiting Low, Moderate and High Degrees of Clustering in Free Recall

Component of ARC Score	Degree of Clustering		
	High	Moderate	Low
Recall protocol	444111333222	123444112233	124334311224
Total number of items recalled (N)	12	12	12
Number of categories represented (k)	4	4	4
Total number of category repetitions (R)	8	5	3
Maximum possible number of category repetitions (maxR)	8	8	8
Expected (chance) number of category repetitions (E(R))	2.0	2.0	2.0
ARC score	1.0	0.50	0.16

Note: In the recall protocols, individual numbers (1, 2, 3, etc.) represent individual items, and repetitions of the same number represent repetition of different items from the same category.

Recall Efficiency

The number of items recalled in each recall task was also of interest as an indicator of the success of strategy training. We anticipated that learning and applying clustering strategies would allow subjects to remember more of the items they had seen or read.

A simple count of the number of items mentioned would not serve as a good measure of recall for a number of reasons. First, the number of items in the memorization tasks used in the time1, time2 and time3 data collections differed. Thus, remembering 7 items at one testing session does not represent the same recall accuracy as remembering 7 items at another testing session. Second, subjects sometimes repeated the same item several times in the course of a recall task. Perhaps an item would be mentioned first because of its primacy or recency in the card deck or paragraph, and would be mentioned again when the subject recalled other items from the same category. In any event, mentioning "bread" twice should not count as remembering two items. Therefore the recall measure could not be based on the literal number of items recalled. Finally, subjects sometimes "recalled" items that were not actually pictured on the cards or included in the prose paragraphs. The spontaneous intrusion of such "unofficial items" increased the variability of the recall protocols and made them less comparable across subjects.

We addressed these issues by using a proportion score to represent accuracy of recall. For each subject we calculated the number of items recalled as a proportion of the number of items that were "available for memorization." The number of items counted as available for memorization depended on whether the stimulus material was picture cards or prose, and whether unofficial items were noted prior to the recall task. In the case of picture recall tasks, we regarded unofficial items as random productions and eliminated them from the count of items recalled. In the case of the paragraphs, however, the unofficial items were sometimes included in the study list which the subject was asked to compose prior to the recall task. In those cases, we regarded the unofficial items as "available for memorization" and counted them if they were in fact subsequently recalled.

Prior to data analysis the proportion scores were corrected using the arcsine transformation (Cohen & Cohen, 1975).

Hypotheses

The strategy training curriculum was designed to introduce the poor readers in this study to the use of categorization as a technique for improving their ability to remember and understand what they read. The technique of identifying groups of items with a common theme, and then labeling those groups, was introduced using pictures of items that the child could actually manipulate and cluster as he or she saw fit. The child was exposed to progressively more difficult types of material. These included (a) sets of pictures which the child was not allowed to physically rearrange, and therefore allowed only mental grouping; (b) prose passages in which all of the items in a group were mentioned together and a category label was supplied; (c) prose passages in which similar items were scattered and the child was required to supply the category label. Thus the child gradually learned to apply the new learning tools to material which resembled normal prose. Each subject was required to achieve a criterion level of categorization and memorization before passing on to the next level of material. In addition, at each step in the process, the tutor would review the categorization and memorization technique using material from the prior stage of the curriculum. This strategy of moving one step back, then one or two steps forward during each tutoring session was intended to consolidate the child's learning as he or she gradually approached the goal of using the categorization strategy on normal prose. The curriculum is described in detail in Appendix B.

We attempted to answer three general questions regarding the success of the strategy training curriculum:

1. Was there an increase in the use of categorization at the most basic level of the curriculum--namely, with regard to the pictures? In other words, did the children in fact learn how to categorize? And if they did learn to categorize the pictures, did this affect their ability to remember what they had seen?
2. If the subjects indeed learned how to categorize pictures, does this learning generalize to other material--specifically, to prose? Was there an increase in the use of categorization with regard to written paragraphs? If so, was there a corresponding increase in the students' ability to remember what they had read?
3. How is the learning and generalization of these skills affected by the number of meetings between student and tutor? Is the use of categorization with picture and

prose material improved by having more training sessions--8 instead of 4, for example? Or is the lesson essentially learned after a few meetings, with minimal or no benefit to be derived from further sessions?

4. How is the generalization of the categorization strategy to prose affected by the type of materials used in the training curriculum? Is the transfer of learning encouraged by the use of prose selections from the child's classroom readings during the training process?

To guide the analysis of the data from the strategy training project, these general questions were translated into the following specific hypotheses:

1. The ARC (categorization) scores based on the free recall of the picture stimuli will (a) be higher for those groups of subjects that have been exposed to the strategy training curriculum than for the control group; and (b) the scores of those subjects in the two long-term treatment groups will be higher following exposure to the curriculum than those of the subjects in the short-term treatment group.
2. The ARC (categorization) scores based on the free recall of the prose stimuli (a) will be higher for the subjects who have been exposed to the strategy training curriculum than for those who have not; (b) the scores of those subjects in the two long-term treatment groups will be higher following exposure to strategy training than the scores of the subjects in the short-term group; and (c) the scores of those subject in the group which

incorporated some classroom-based materials in the training process will exhibit higher use of categorization (ARC scores) than subjects in the other long-term treatment group.

3. The memorization scores based on the free recall of the picture stimuli (a) will be higher among those subjects that have been exposed to the strategy training curriculum than among those who have not; and (b) the mean memorization score for those subjects in the two long-term treatment groups will be higher following exposure to the curriculum than that of the subjects in the short-term treatment group.
4. The mean memorization score based on the free recall of prose (a) will be higher for those groups that have been exposed to strategy training than it will be for the control group; (b) will be higher for the two long-term treatment groups than for the short-term treatment group; and (c) will be higher for that long-term group which included classroom-based materials in the curriculum than for the long-term group which used only the specially prepared prose examples.

The specific statistical analyses used to test each hypothesis, and the results, are discussed in the next section.

RESULTS

Overview

A basic premise of the strategy training curriculum was that the subject would find it easier to grasp the notion of categorization if it were applied first to concrete objects. Therefore the curriculum consisted of a graduated series of tasks, beginning with the categorization of pictures of objects such as a motorcycle, a chair, and an airplane, and gradually worked toward the application of the categorization strategy to an excerpt from a history textbook. The first question addressed by the data analysis concerned this basic assumption: did the children in fact grasp the notion of categorization when this was presented in its simplest, most straightforward form?

The next question which we addressed concerned the application of the categorization strategy to written material. If the subjects did indeed understand the notion of categorization, as witnessed by their performance with pictures, was the curriculum successful in promoting the transfer of that cognitive strategy to another medium--namely, prose?

A second basic assumption of the strategy training curriculum was that the use of categorization would improve the ability of the poor readers in the study to retain what they had seen or read. In other words, increased categorization would be linked to increased recall efficiency. The relationship between categorization and recall was the third issue addressed in the analyses described below.

Finally, a distinction can be made between those subjects who did grasp the idea of categorizing, and those who did not. Seven of the 47 children in this study fell into the latter group. Although these children did not differ from the other subjects with regard to either IQ or digit span recall capacity, their poor performance on the picture categorization and recall tasks suggested that they did not understand that categorization was a cognitive tool in the same manner as did the more insightful of the poor readers in the study. The differences between the "insightful" and "uninsightful" children with regard to the categorization and recall of pictures and prose were addressed by the final series of analyses reported here.

In summary, the following questions were addressed:

(1) Did the subjects learn to categorize picture cues?

(2) Having learned to categorize pictures, were the subjects subsequently able to apply this strategy to written material in paragraph form?

(3) Did the use of categorization increase the subjects' ability to remember what they had seen (pictures) or read (prose)?

(4) Were there children who seemed unable to grasp the basic insight of the strategy training curriculum? If so, how did their performance differ from that of the more "insightful" children?

Before proceeding to a detailed discussion of the formal statistical analyses, it would be helpful to review the design of the experiment. Figure 1 summarizes the relationship between the testing sessions (pre-test and two posttests) and the actual conduct of the strategy training tutorials in the three treatment groups. Also indicated is the sequence of picture and prose tasks used to assess use of the categorization strategy and recall efficiency.

Figure 1

Schematic Summary of Experimental
Design and Data Collection:
Strategy Training Project

Group		Data collection				
		Time 1 (pretest)		Time 2 (Posttest 1)		Time 3 (Posttest 2)
Piles		1A 2A 2B * 1B		3 4A 4B		5 6A 6B
Paragraphs		1 2		3 4		5
Control	(12)	0		0		0
Short-term treatment	(11)	0	two weeks of tutorials	0		0
Long-term 1	(11)	0	two weeks of tutorials	0	two weeks of tutorials	0
Long-term 2	(13)	0	tutorials (2 weeks)	0	tutorials (2 weeks)	0

Categorization of Pictures

As previously described, all subjects were tested prior to the beginning of strategy training to establish their baseline rate of spontaneous categorization and recall efficiency. As their first task in the pre-test session, subjects were asked to arrange a deck of 15 cards "in one pile in a way that will make it easy for you to remember the pictures." Subjects were allowed to review their piles, and then asked to recall as many of the items as they could. At the conclusion of the pretest session, the tester demonstrated with different picture stimuli that the pictures could be sorted into groups of items that shared a common theme (e.g., items of furniture). Each subject was then given an opportunity to rearrange his/her deck so as to take advantage of the grouping strategy. Subjects were then asked to again recall as many of the items as they could.

To determine whether the children understood and applied the grouping strategy, the extent of categorization evident in the items recalled _after the demonstration_ was compared with the level of categorization evident in the first trial, _prior to the demonstration_ of the "hint" of grouping items "that go together." As discussed in the METHODS chapter above, the adjusted ratio of categorization (ARC score) was used to measure the extent of categorization. Paired T-tests were used to compare the pre-demonstration and post-demonstration levels of categorization, for all subjects and for each treatment group.* The results are reported in Table 6. In all cases, the post-demonstration level of categorization was significantly higher than the pre-demonstration level.

Each tutorial in the strategy training program began with an exercise involving pictures. This routine continued even after the focus of the tutorials shifted from the introduction of the concept of grouping using pictures (which most subjects grasped by the end of the first meeting) to the application of grouping strategy to prose. The purpose of this regimen was to repeatedly reinforce the concept of

---

*Due to the limitations of the programming language used, the actual procedure employed here was to create difference scores (post-demonstration ARC score minus pre-demonstration ARC score) and test whether the mean of these change scores was significantly different from zero.

TABLE 6

Extent of Categorization Evident in Recall
Before and After Demonstration of "Grouping":
Mean ARC Scores for All Subjects and for Treatment Groups

Group (N)	Trial A (Before)	Trial B (After)	Mean Difference	T
All Subjects (32)	.54	.87	.33	4.48***
Control Group (7)	.40	.79	.39	2.02*
Short Term Treatment (8)	.78	.95	.17	2.02*
Long Term Treatment 1 (9)	.61	.87	.26	3.16**
Long Term Treatment 2 (8)	.34	.84	.50	2.47*

concept of categorization by returning to the most concrete examples of grouping--to "touch base" with the categorization of pictures of objects at the beginning of each lesson before moving on to exercises involving the categorization of items mentioned in paragraphs.

Similarly, each testing session included picture tasks as well as prose tasks. Thus it was possible to determine whether the increased use of categorization observed immediately following the demonstration at the end of the pre-test session was sustained for the duration of the project. That is, did the subjects retain and continue to apply the strategy of categorization for several weeks after the technique of "grouping" was introduced to them?

Paired t-tests were used to examine this question. The extent of categorization evident in each subject's recall protocol for the first trial of the first pile (i.e., prior to the demonstration of the "grouping" hint) was again used as the base rate of categorization. This was compared with the extent of categorization evident in recall of pictures at (a) posttest 1, after two weeks of strategy training, and (b) at posttest 2, after two additional weeks of instruction.

As described above, subjects were allowed to manually sort some piles of pictures into the order they preferred (subject-sorted piles). For other piles, subjects were allowed only to study the pictures in the order that they were presented (fixed-order piles). Rearranging the pictures in these fixed-order piles was not permitted; any sorting done was "mental" sorting. The manual sorting instructions allowed for the subject to manipulate the picture stimuli, and thus spatially represent the categories he/she used in sorting these stimuli. It was thus expected that the categorization strategy would be more consistently used and would improve recall efficiency to a greater degree in "subject-sorted" piles than in "fixed-order" piles of picture stimuli.

The results are reported in Table 7. The comparisons between the pre-demonstration levels of categorization and both posttests were significant in the direction expected. The increase in the use of categorization evident in the exercise immediately following the grouping demonstration was sustained over the duration of the strategy training program. This was true for both subject-sorted and fixed-order piles, although in general the subjects performed much better with the subject-sorted piles.

TABLE 7

Extent of Categorization Evident in Recall
at Posttest 1 and Posttest 2:
Mean ARC Scores for All Subjects by Type of Pile

Time of Testing	Type of Pile Subject-Sorted	 Fixed-Order
Pre-Test	.66	.15
PostTest 1	.93	.33
T (Pre to Post 1)	4.37*** (n=43)	2.18** (n=40)
Pre-Test	.64	.15
Posttest 2	.89	.46
T (Pre to Post 2)	3.50*** (n=31)	3.59*** (n=34)

Note: ARC scores for pretest differ slightly for each comparison (Post 1, Post 2) and for sorted vs. fixed piles because number of subjects with valid recall data for the piles in question varies due to attrition at later testing times, audio tape failure, and so on.

** p<.01, one-tailed.

*** p<.001, one tailed.

The category instruction in the pre-test for all subjects, then, significantly improved all subjects' ability to categorize picture stimuli at post-test. Since the categorization strategy was the central focus of the training in this study, systematic investigation of the effects of the treatment conditions on subjects' ability to use the categorization strategy was required. In the analysis of treatment effects, three planned comparisons were used to illuminate various aspects of the process of learning and applying the strategy of categorization to picture and prose stimuli. These comparisons were used in all subsequent analyses. The three planned comparisons were: (1) control vs. all treatment conditions, to investigate the acquisition of the strategy trained; (2) control vs. short-term treatment conditions, to investigate retention of the trained strategy; and (3) short-term vs. long-term treatment conditions, to investigate the amount of training required to produce effective strategy acquisition and retention. All of these questions were addressed using analysis of covariance with specified a-priori contrasts (Bock, 1975). These comparisons are summarized in Table 8.

Control vs. All Treatment Groups: Acquisition

To answer the basic question of whether the strategy training tutorials had had any effect on the use of categoriztion, the three treatment groups (combined) were compared with the control group at Time 2 (posttest 1) using Time 1 (pretest) level of categorization as a covariate. This was done separately for the subject-sorted and fixed-order piles. The results are presented in Table 9 (Row 1). After controlling for pre-test level of categorization, the difference between the control group and the treatment groups was no greater than one would expect by chance. This was true both for the subject-sorted and fixed-order piles.

Table 8

Overview of Planned Contrasts Used to Evaluate

the Categorization of Picture Stimuli:

Analysis of Covariance Model

Effect Tested	Groups	Picture Stimuli (Piles)			
		Subject-Sorted		Fixed-Order	
		DV	COV	DV	COV
No training vs. training re: aquisition of categorization strategy	Control vs. all treatment groups	Pile 3 (Time 2)	Pile 1A (Time 1)	Pile 4A (Time 2)	Pile 2A (Time 1)
No training vs. training re: retention of categorization strategy	Control vs. short-term treatment at Time 3	Pile 5 (Time 3)	Pile 1A (Time 1)	Pile 6A (Time 3)	Pile 2A (Time 1)
Some training vs. extended training re: acquistion of categorization strategy	Short-term vs. both long-term treatment groups, at completion of curriculum	Short term: Pile 3 (Time 2)	Pile 1A (Time 1)	Pile 4A (Time 2)	Pile 2A (Time 1)
		Long term: Pile 5 (Time 3)	Pile 1A (Time 1)	Pile 6 (Time 3)	Pile 2A (Time 1)

DV = dependent variable
COV = covariate

Table 9

Extent of Categorization Evident in Recall of Pictured Items: Group Contrasts

Contrast Groups	Mean ARC Scores							
	Subject-Sorted				Fixed-Order			
	Pre	(N)	Post	F+ (df)	Pre	(N)	Post	F+ (df)
Control vs. All Treatment, Time 2	.66	(10)	.98	<0.50	.20	(10)	.38	<1.00
	.66	(33)	.92	(4,38)	.13	(30)	.31	(4,35)
Control vs. Short Term Treatment, Time 3	.78	(8)	.84	1.00	.18	(8)	.55	1.46
	.86	(4)	.84	(4,26)	.29	(6)	.30	(4,29)
Short Term vs. Long-Term, at Conclusion of Training	.71	(11)	.93	<0.50	.09	(11)	.21	2.74*
	.54	(19)	.93	(2,27)	.25	(20)	.46	(2,28)

+ F for overall model

*p<.10

Control vs. Short Term Treatment Group: Retention

A distinction can be made between the acquisition of the categorization strategy and the retention of that strategy. As discussed above, the use of categorization evident in the sequence of items recalled by all subjects showed an increase following the demonstration of grouping with picture stimuli which was given at the end of the pretest session. The subsequent two weeks of tutorials did not substantially increase the level of categorization among the treatment groups. However, does the treatment make a difference in the retention and maintenance of the cognitive strategy over time? To answer this question, the short-term treatment group and the control group were compared at time 3, two weeks after the short-term treatment subjects had completed the strategy training curriculum.* It was predicted that participation in the strategy training project would result in a higher use of categorization at time 3 than would be evident in the protocols of the control group children, who had only been exposed to the single pre-test session demonstration.

Again an analysis of covariance was used to compare the time 3 ARC scores of the control and short-term treatment subjects, controlling for baseline level of performance (time 1 ARC score). The results are reported in Table 9 (Row 2). The difference between the mean ARC scores at Time 3 was not statistically significant.

---

*At time 3 the long-term treatement groups had just completed the curriculum. Thus, a fourth data collection, two weeks later, would have been necessary to assess the attentuation of the categorization strategy in those groups. By contrast, at time 3 the short-term group had gone two weeks since their last tutorial.

Short-term vs. Long-term Treatment Groups

The third hypothesis to be examined concerned the distinction between the long-term and the short-term versions of the strategy training program. Did the children assigned to the long-term treatment conditions demonstrate a greater tendency to use categorization at the conclusion of their training (4 weeks) than did the short-term treatment subjects at the conclusion of theirs (2 weeks)? Again the analysis of covariance was used, with the baseline rate of categorization a covariate. The indicator of post-treatment categorization was the ARC score obtained at the testing session immediately following the conclusion of training for the two groups. For the short-term treatment subjects, this was time 2 testing; for the long-term treatment group, the time 3 score was used. The results of these analyses (Table 9, Row 3) demonstrate that differences between these two groups approached significance in the fixed-order piles ($p < .10$).

Categorization of Prose

In general, the subjects in the current study were able to grasp the notion of categorization as demonstrated using pictures of concrete objects. Furthermore, the subjects were able to use this cognitive strategy whenever they were required to recall piles of pictures throughout the four week period of the strategy training period. Thus, the insight gained as a result of the demonstration at the end of the pre-test session, and subsequently reinforced for the children in the treatment conditions, was retained by most of the subjects. However, was the strategy training curriculum successful in promoting the transfer of this skill from pictures to prose?

Several subsidiary questions are involved in this issue of the transfer of learning in the strategy training project. In addition to the basic question of whether the subjects in the treatment groups performed better than those in the control condition, both 2 week and 4 week versions of the curriculum were used in order to permit exploration of the question of "How much was enough?" Furthermore, the prose passages were designed to gradually approximate textbook prose so as to facilitate the transition from pictures to objects to prose in which the "objects" were "ideas." In some passages, the items from each category were mentioned together in one or two sequential sentences (the "blocked" paragraphs). Other passages required more effort in that categories could only be formed by grouping items which shared a common theme but which were mentioned in different sentences scattered throughout the passage (the "unblocked" paragraph). Thus, in addition to the contrasts between no treatment and treatment and between short term and long term treatment, there was also the dimension of "blocking" to take into consideration in the analysis of the data from the prose passages. Perhaps the subjects were able to make the transition from picture to prose only when the prose material was blocked. Perhaps the transfer to unblocked prose was accomplished only by those subjects in the long-term treatment groups.

All of these questions were addressed using analysis of covariance with specified a-priori contrasts (Bock, 1975). Specifically, we identified four combinations of treatment condition and prose structure which represented those evaluations of the effect of the strategy training program which were possible within the limitations of the data available. These are summarized in Table 10.

Table 10

Overview of Planned Contrasts Used to Evaluate the Transfer of Categorization to Prose Material: Analysis of Covariance Model

Effect Tested	Groups	Type of Prose			
		Blocked		Unblocked	
		DV	COV	DV	COV
No training vs. training: aquisition of categorization strategy	Control vs. all treatment groups	Par 3 (Time 2)	Par 2 (Time 1)	Par 4 (Time 2)	Par 1 (Time 1)
No training vs. training: retention of categorization strategy	Control vs. short term treatment at Time 3	Blocked data not available for this contrast		Par 5 (Time 3)	Par 1 (Time 1)
Some training vs. extended training re: acquistion of categorization strategy	Short Term vs. both long-term treatment groups, at completion of curriculum	Blocked data not available for this contrast		Short term: Par 4 (Time 2) Long term: Par 5 (Time 3)	 Par 1 (Time 1) Par 1 (Time 1)

DV = dependent variable
COV = covariate

Control vs. All Treatment Groups: Acquisition

To answer the basic question of whether the strategy training tutorials had any effect on the use of categorizaton in prose, the three treatment groups (combined) were compared with the control group at Time 2 (posttest 1) using Time 1 (pretest) level of categorization in prose as a covariate. This was done separately for the blocked and unblocked passages. The results are presented in Table 3-6. After controlling for pre-test level of categorization, the difference between the control group and the treatment groups was no greater than one would expect by chance. This was true both for the unblocked and the blocked material.

Control vs. Short-term Treatment Group: Retention

As with the categorization of picture stimuli, after two weeks of training there were no significant differences between treatment and control groups with respect to the categorization in prose measures. However, the gains in categorization in prose made by the control group may have been due to the power of category instruction given to all groups during the pre-test session. Therefore, the gains in categorization in prose may not have been retained over time by the control group. It was predicted that participation in training sessions would result in higher use of categorization in prose at time 3 by subjects in the short-term training condition than would be evident in the protocols of the control children, who had only been exposed to a single pre-test session demonstration.

Again an analysis of covariance was used to compare the time 3 ARC scores of the control and short-term treatment subjects, controlling for baseline level of performance (time 1 ARC score). The results are reported in Table 12. Note that only data for unblocked prose passages were available for this contrast.

TABLE 11

Categorization Evident in Recall of Items From Blocked and Unblocked Prose Passages at Time 1 and Time 2: Mean ARC Scores

Group	Type of Prose					
	Blocked			Unblocked		
	Pre	(N)	Post	Pre	(N)	Post
Control	.57	(9)	.82	.60	(9)	.60
Short Term	.48	(10)	.85	.24	(11)	.74
Long Term 1	.41	(9)	.84	.48	(9)	.59
Long Term 2	.67	(13)	.80	.41	(13)	.75

Blocked

Overall Model	F (4,36) = 1.62	p>.18
Treatment Contrast	F (1,36) = 0.04	p>.50

Unblocked

Overall Model	F (4,37) = 1.30	p>.25
Treatment Contrast	F (1,37) = 1.25	p>.25

TABLE 12

Categorization Evident in Recall of Items From Prose Passages at Time 3: Mean ARC Score Values

Group (N)	Categorization at Time 1	 Time 2
Control (6)	.47	.70
Short-term Treatment (4)	.30	.57

Overall Model: F (4,25) = 0.35 p>.50

Treatment Contrast: (Control vs. Short-term) F (1,25) = 0.46 p>.50

The difference between the mean ARC scores at Time 3 was not statistically significant ($F_{1,25} = 0.46$; $p>.50$).

Short-term vs. Long-term Treatment Groups

The third hypothesis to be examined concerned the distinction between the long-term and the short-term versions of the strategy training program. Did the children assigned to the long-term treatment conditions demonstrate a greater tendency to use categorization in prose at the conclusion of their training (4 weeks) than did the short-term treatment subjects at the conclusion of theirs (2 weeks)? Again the analysis of covariance was used, with the baseline rate of categorization in prose a covariate. The indicator of post-treatment categorization was the ARC score obtained at the testing session immediately following the conclusion of training for the two groups. For the short-term treatment subjects, this was time 2 testing; for the long-term treatment group, the time 3 score was used. Only data based on unblocked prose was available for this contrast.

The results are reported in Table 13. The short-term subjects displayed a higher mean level of categorization (ARC score .74) than did the children in the long term treatment groups (ARC score .60), but the difference was not statistically significant.

TABLE 13

Categorization Evident in Recall of Items From Prose Passages at Conclusion of Training: Short Term vs. Long Term Treatment

Group (N)	Categorization (ARC Score) Time 1	Post-training
Short-term Treatment (11)	.24	.74
Long-term Treatment groups (20) (combined)	.36	.60

Overall Model: F (2,28) = 0.97 p>.25

Treatment Effect: F (1,28) = 1.33 p>.25

Relationship Between Categorization and Recall Efficiency

Recall Efficiency

Recall of pictures. The proportion of pictured items recalled from each pile at time 1, time 2 and time 3 is reported in Table 14. The data is summarized for all subjects and for each experimental group. There was a general increase over time in the number of items recalled. This pattern holds for all groups, controls as well as the three treatment groups. Furthermore, unlike the findings with regard to extent of categorization, the trend was found to have the same direction and order of magnitude regardless of whether the piles were subject-sorted or fixed-order.

To formally examine the question of differences in recall efficiency between the groups, a series of analyses of covariance were conducted. It was predicted that subjects who participated in the

Table 14

Proportion of Pictured Items Recalled

Group	Time, Pile Number, and Pile Type									
	Subject-Sorted				Fixed-Order					
	Time 1		Time 2	Time 3	Time 1		Time 2		Time 3	
	1A	1B	3	5	2A	2B	4A	4B	6A	6B
All Subjects	64 (47)	82 (32)	77 43)	80 (32)	58 (47)	77 (47)	68 (43)	86 (40)	61 (33)	80 (34)
Control	63 (12)	80 (7)	78 (10)	78 (8)	58 (12)	78 (12)	68 (10)	81 (10)	63 (8)	81 (8)
Short Term	62 (11)	88 (8)	76 (11)	68 (5)	61 (11)	80 (11)	64 (11)	89 (11)	52 (5)	69 (6)
Long Term 1	60 (11)	78 (9)	81 (9)	79 (9)	56 (11)	73 (11)	69 (9)	84 (8)	56 (9)	76 (9)
Long Term 2	71 (13)	81 (8)	75 (13)	89 (10)	58 (13)	78 (13)	71 (13)	89 (11)	70 (11)	89 (11)

Note: A and B refer to the first and second trials respectively of the same pile.

strategy training project would demonstrate greater recall efficiency than the subjects in the no-treatment control group. Furthermore, it was anticipated that the long-term treatment groups would perform better at the conclusion of their participation in the project than would the short-term treatment group. A series of planned contrasts, parallel to those used to evaluate the impact of strategy training on the use of categorization, were performed (see Table 8). In all cases, the baseline rate of recall efficiency (time 1 testing) was used as a covariate. Recall data from subject-sorted and fixed-order piles were evaluated in separate analyses.

Three of the six analyses resulted in statistically significant F ratios for the general model tested in each case (see Table 15). For both subject-sorted and fixed-order piles, the short-term and long-term treatment groups differed to an extent greater than would be expected by chance when each group was tested immediately upon completing the strategy training curriculum. However, the direction the results was inconsistent. With regard to subject-sorted piles, the main effect for the treatment group was significant, F (1,27) = 3.86, p = 0.59. The long-term treatment group (mean percent of items recalled = 84) outperformed the short-term treatment group (mean = 76%). Recall of items from fixed-order piles exhibited a significant treatment main effect, F (1,28) = 127.96, p <.001. However, the group means displayed the opposite pattern; the short-term treatment subjects (m = 89%) remembered a higher proportion of the pictures they saw than did long-term treatment subjects (m = 83%).

The other significant result was obtained when contrasting the control-group and the short-term treatment group at time 3 with regard to their recall of items from fixed-order piles (see Table 15). However, the a-priori contrast between these two groups was not the source for the signifcant overall F ratio for the model (F 1,29 = 0.05; p >.50). A significant main effect for treatment was observed. Post-hoc tests between the means of the four treatment groups (Duncan multiple range tests) subsequently indicated that the mean recall proportion for subjects in the long-term 2 treatment group (m = 89%) was significantly different from either the long-term 1 treatment mean (76%) or the short-term treatment mean (69%).

Table 15

Proportion of Pictured Items Recalled

Contrast Groups	Mean Proportion Recalled							
	Subject-Sorted				Fixed-Order			
	Pre	(N)	Post	F+ (df)	Pre	(N)	Post	F+ (df)
Control vs. All Treatment Groups at Time 2	65	(10)	78	<1.0	83	(10)	81	1.74
	65	(33)	77	(4,38)	76	(30)	86	(4,35)
Control vs. Short-Term Treatment at Time 3	72	(8)	78	1.85	85	(8)	81	4.28**
	61	(5)	68	(4,27)	86	(6)	69	(4,29)
Short Term vs. Both Long-Term Groups at Conclusion of Training	62	(11)	76	3.56*	80	(11)	89	68.89***
	64	(19)	84	(2,27)	74	(20)	83	(2,28)

Note: Simple proportion of items recalled are reported above for ease of interpretation. ARC sine transformed proportions were used in the actual analysis of covariance calculations (Cohen & Cohen, 1975). Decimal point omitted in tabled values.

+ F for overall model

* p<.05 ** p<.01 ***p<.001

Recall of Prose. The proportions of items recalled from the prose passages used in the time 1, time 2 and time 3 assessments of the subjects' recall efficiency are reported in Table 16. The data are summarized for all subjects and for each group. In general, all subjects demonstrated a relatively high baseline rate of item recall that was not substantially lower for the unblocked passages than it was for the blocked passages. Over time, most groups tended to either improve their recall rate slightly or to remain at roughly the same level. There were some exceptions to these patterns. With regard to unblocked prose, the recall efficiency of the control group declined over time. Also with regard to the unblocked prose, the short-term treatment group started at a lower average level of recall (m = 76%) than all other groups, demonstrated some improvement at time 2 (m = 79%), but ended up a lower level of recall at time 3 (73%).

Again, a series of analyses of covariance incorporating a-priori planned contrasts between various groups were used to formally test hypotheses related to the impact of the strategy training curriculum. It was predicted that subjects in the treatment groups would demonstrate greater recall efficiency than subjects in the control group at time 2 (i.e., after the participants had attended their first two weeks of tutorials). It was predicted that participation in the strategy training program would lead to a more stable improvement in recall efficiency than would the experience of the single pretest session. This hypothesis was tested by comparing the recall of the short-term treatment group and the control group at time 3 (two weeks after the conclusion of instruction for the short-term group). As before, the third major hypothesis was that more strategy training would result in greater recall efficiency than would less training. This was tested by comparing the recall scores of the short-term group

Table 16

Proportion of Items Recalled from Prose Passages

Group	Blocked Paragraphs		Unblocked Paragraphs		
	Time 1	Time 2	Time 1	Time 2	Time 3
All Subjects	86 (46)	90 (44)	81 (46)	85 (43)	81 (31)
Control Groups	87 (11)	87 (11)	83 (11)	86 (10)	71 (7)
Short-Term Treatment	84 (11)	88 (11)	76 (11)	79 (11)	73 (4)
Long-Term Treatment 1	86 (11)	86 (9)	85 (11)	90 (9)	85 (9)
Long-Term Treatment 2	85 (13)	96 (13)	81 (13)	86 (13)	86 (11)

Paragraph Type, Time, and Paragraph Sequence

Note: Blocked material not available at Time 3.

obtained at the completion of training (time 2) with the recall scores of the long-term treatmentt groups obtained at the conclusion of their training (time 3). In all cases, the baseline rate of recall efficiency (time 1 testing) was used as a covariate. Data from blocked and unblocked assessment tasks were analyzed separately.

The results are summarized in Table 17. Four analyses were conducted. Blocked material was not included in the time 3 testing routine, thus eliminating two of the six possible a-priori contrast analyses. Only one overall F-ratio achieved statistical significance, that involving recall from unblocked prose at time 2. However, this was due neither to a significant main effect for treatment group in general or to the planned contrast between the control group and all treatment groups. Rather, the pre-test differences in recall efficiency included as a covariate were responsible for the greater than chance amount of variance accounted for by the model as a whole ($F_{1,37} = 11.91$; $p < .001$).

Table 17

Proportion of Items Recalled from Prose Passages

Contrast Groups	Mean Proportion Items Recalled							
	Blocked				Unblocked			
	Pre	(N)	Post	F+ (df)	Pre	(N)	Post	F+ (df)
Control vs. All Treatment, Time 2	87	(10)	88	2.08	86	(9)	89	3.75**
	85	(33)	91	(4,38)	85	(33)	91	(4,37)
Control vs. Short Term, Time 3	Data not available				93	(6)	75	1.00
					74	(4)	73	(4,25)
Short Term vs. Long-Term, at Conclusion of Training	Data not available				76	(11)	88	1.70
					82	(20)	86	(2,28)

Note: F ratios reported above are for overall model in question (i.e., for treatment condition and covariance as predictors of dependent variable)

Categorization and Recall

To directly assess the relationship between categorization and recall, Pearson correlation coefficients were calculated between the ARC scores (indicative of extent of categorization) and the arcsine transformed proportion of items recalled for each pile and each paragraph. This was done for the sample as a whole, and separately for each group.

The results with regard to pictures are reported in Table 18. As before, a distinction has been made between subject-sorted and fixed-order piles. No simple generalizations appear appropriate in either case. The relationship between the use of categorization and recall efficiency when subject-sorted piles were involved was weak but positive across all subjects. This "average" relationship represents contradictory tendencies within experimental groups. Within the short term group and one long term group, the relationship between use of categorization and recall efficiency became increasingly positive over time. The control group displayed a decrease; the relationship within the second long term treatment group was erratic. In general, with regard to fixed-order piles, the strength of the tie between use of categorization and efficiency of recall decreased over time. The second long term treatment group demonstrated the opposite trend.

The categorization-recall relationship with regard to items presented in prose passages is summarized in Table 19. Separate calculations were computed for blocked and unblocked paragraphs. Among the former, categorization displayed a moderate to strong positive relationship with recall, and this generally improved over time. With regard to unblocked passages, a moderately strong and stable relationship between use of categorization and recall efficiency was discovered. However, this varied markedly from

Table 18

Correlation Between Categorization and Recall - Pictures

Group	Subject-Sorted			Fixed-Order		
Time	1	2	3	1	2	3
All Subjects	.17 (32)	.25 (43)	.12 (31)	.30 (47)	.16 (40)	.13 (34)
Control	.26 (7)	.14 (10)	.14 (8)	.34 (12)	-.12 (10)	.17 (8)
Short-Term Treatment	-.37 (8)	.48 (11)	.52 (4)	.53 (11)	.58 (11)	.19 (16)
Long-Term 1	.28 (9)	.13 (9)	.39 (9)	.13 (11)	.44 (8)	.04 (9)
Long-Term 2	.08 (8)	.36 (13)	-.57 (10)	.18 (13)	.07 (11)	.44 (11)

Note: Values reported are Pearson correlation coefficients between ARC scores (categorization) and arcsine transformed proportion of items recalled (recall efficiency) for the same pile (e.g., the fixed-order pile used in the time 2 data collection). Also, all fixed-order piles were presented twice; the correlations reported above are based on categorization and recall of the second trial at each testing.

Table 19

Correlation Between Extent of Categorization and Recall Efficiency - Prose

Group	Type of Paragraph and Time				
	Blocked Paragraphs		Unblocked Paragraphs		
Time	1	2	1	2	3
All Subjects	.26 (45)	.41** (43)	.30* (46)	.15 (43)	.28 (31)
Control	.34 (11)	.67* (11)	.27 (11)	.03 (10)	.10 (7)
Short-Term	.06 (10)	.49 (11)	.60* (11)	.02 (11)	.63 (4)
Long-Term 1	.45 (11)	.24 (9)	-.40 (11)	.82** (9)	.62 (9)
Long-Term 2	.17 (13)	.27 (13)	.38 (13)	.03 (13)	.10 (11)

Note: Values reported are Pearson correlation coefficients between ARC score (categorization) and arcsine transformed proportion of items recalled (recall efficiency) for the same prose passage (e.g., the unblocked passage used in the time 2 data collection.

* p <.05 ** p <.01

group to group. Within the short-term treatment group, the relationship was strong and stable (r = .60 at time 1 and .63 at time 3). Within both the control group and the second long term group, the relationship decreased from a moderate level to a low level over time. The first long-term treatment group displayed a marked increase in the strength of the relationship between categorization and recall between time 1 and time 3.

Insightful and Uninsightful Children

Definition and Distribution

The notion of categorization was presented to the subjects as a "trick" to improve one's ability to remember pictured items in the very first tutorial. Recall performance criteria were set for each task. For example, subjects did not graduate to the next, more difficult step in the training process until they had correctly recalled 3 out of 4 items in each of four out of a possible five categories represented in the first picture deck. Several "back-up" tasks were included in the curriculum in the event that some of the students did not understand and effectively apply the categorization strategy following its initial presentation. For example, the initial "back-up" task for the first training deck of pictures (20 items) was to simply use a smaller deck (8 items, four in each of two categories). If the child was still unable to recall the criterion number of items, appropriately clustered, a second and even third "back-up" task was available which further simplified the task and reinforced the use of categorization (for example, by physically grouping together the pictures of those items which belonged to the same category). At whatever level of back-up task the child did reach the recall criterion, the subject was then required to work his or her way back up through the sequence of back-up tasks successfully, and to then ultimately continue with the primary series of tasks.

For the majority of subjects, the back-up tasks were not required. However, 7 of the 35 subjects in the three treatment groups (20%) did need one or more of the additional, simplified tasks before they were able to categorize to the criterion level and continue with the primary sequence of tasks in the curriculum.* The distribution of these "uninsightful" children across the three treatment conditions is reported in Table 20. The uninsightful children did not differ from their insightful peers with regard to either IQ (t (44) = -1.06; p>.14) or baseline recall capacity, as assessed by the WISC digit-span task (T (45) = -0.34; p>.10); see Table 21.

There were, however, consistent performance differences between the two groups. The mean level of categorization (ARC scores) for uninsightful and insightful children are reported in Table 22 for both pictured items and items mentioned in prose. Formal statistical comparisons were conducted for fixed-order and subject-sorted piles separately, and likewise for blocked and unblocked paragraphs. In all but a few tasks, less category clustering was evident in the recall protocols of the unsightful children; in most instances this difference was statistically significant.

---

*Of course, it was not possible to determine how many of the 12 control subjects might also fall into this group because the control group was not exposed to the categorization tasks whereby the "uninsightful" children were identified.

Table 20

Distribution of Uninsightful Children
Among Treatment Groups

Insight Group		Treatment Group		
		Short Term	Long Term 1	Long Term 2
Insightful	N	9	6	13
	%	82%	55%	100%
Uninsightful	N	2	5	0
	%	18%	45%	0%

A similar pattern was observed with regard to recall. The mean proportions of items recalled for insightful and uninsightful children are summarized in Table 23 for both pictures and prose. Again, formal statistical comparisons were conducted separately for subject-sorted and fixed-order piles, and for blocked and unblocked paragraphs. The insightful children recalled more than their uninsightful peers in all but one instance (picture deck 3), and 8 of the 12 comparisons were statistically significant in favor of the insightful group.

Table 21

IQ and Digit-Span Recall Differences

Between Insightful and Uninsightful Children

Characteristic	Group Means			
	Insightful	Uninsightful	t(df)	p[3]
IQ[1]	93.2 (39)	86.3 (7)	-1.07 (44)	.15
Digit-Span Recall[2]	4.7 (40)	4.1 (7)	-1.25 (45)	.11

1. As assessed by Peabody Picture Vocabulary Test.
2. Length of longest forward digit string successfully recalled in two successive trials, WISC digit span recall task. Maximum = 9.
3. One-tailed values; it was hypothesized that the uninsightful children would have lower IQ scores and lower digit-span recall capacity

Table 22

Differences in Use of Categorization Between Insightful and Uninsightful Children: Mean ARC Scores

	Group			
Stimulus Characteristics	Insightful	Uninsightful	t(df)	p (one-tailed)
Pictures				
Subject-sorted Piles				
Deck 1A (Time 1)	.63 (40)	.61 (7)	<0.50 (45)	>.45
1B (Time 2)	.89 (26)	.76 (6)	-1.54 (30)	.06
3 (Time 2)	.96 (37)	.79 (6)	-2.66 (41)	<.01
5 (Time 3)	.90 (25)	.85 (6)	-0.74 (29)	>.20
Fixed-Order Piles[1]			[2]	
Deck 2B (Time 1)	.12 (40)	.20 (7)	0.34 (67)	>.35
4B (Time 2)	.34 (34)	.28 (6)	0.40 (38)	>.30
6B (Time 3)	.43 (28)	.60 (6)	0.90 (32)	>.15
Prose				
Blocked Paragraphs				
Par 2 (Time 1)	.57 (39)	.04 (6)	-2.87 (43)	<.01
Par 3 (Time 2)	.83 (37)	.56 (6)	-1.86 (41)	<.05
Unblocked Paragraphs				
Par 1 (Time 1)	.41 (39)	.34 (7)	-0.44 (44)	>.30
Par 4 (Time 2)	.72 (37)	.40 (6)	-2.23 (41)	<.02
Par 5 (Time 3)	.70 (25)	.31 (6)	-2.72 (29)	<.01

---

1. To simplify the table, only the mean ARC score for the second (B) trial of the same deck of pictures is reported here. In general, performance on the second trial of the same deck was better than the first.

2. Preliminary analysis indicated significant differences in variance between the two groups. Therefore, separate variance estimates were used in calculation of t rate.

Table 23

Differences in Recall Accuracy Between Insightful and Uninsightful Children: Mean Proportion of Items Recalled

	Group			
Stimulus Characteristics	Insightful	Uninsightful	t(df)	p (one-tailed)
Pictures				
Subject-sorted Piles				
Deck 1A (Time 1)	67 (40)	51 (7)	-3.04 (18.8)[2]	<.01
1B (Time 2)	83 (26)	76 (6)	-0.91 (30)	>.15
3 (Time 2)	76 (37)	86 (6)	1.26 (41)	>.10
5 (Time 3)	82 (26)	73 (6)	-0.95 (30)	>.15
Fixed-Order Piles[1]				
Deck 2B (Time 1)	80 (40)	63 (7)	-1.98 (45)	<.05
4B (Time 2)	87 (34)	78 (6)	-1.47 (38)	<.08
6B (Time 3)	84 (28)	64 (6)	-2.79 (32)	<.005
Prose				
Blocked Paragraphs				
Par 2 (Time 1)	87 (39)	71 (7)	-2.63 (44)	<.01
Par 3 (Time 2)	90 (38)	76 (6)	-1.94 (42)	<.05
Unblocked Paragraphs				
Par 1 (Time 1)	82 (39)	79 (7)	-1.86 (32.9)	<.05
Par 4 (Time 2)	86 (37)	76 (6)	-1.68 (41)	<.05
Par 5 (Time 3)	81 (25)	79 (6)	-0.65 (29)	>.25

Note: Raw proportions are reported above for ease of interpretation. Arcsine transformed proportions were used for all formal statistical comparisons.

1. To simplify the table, only the mean ARC score for the second (B) trial of the same deck of pictures is reported here. In general, performance on the second trial of the same deck was better than the first.

2. Preliminary analysis indicated significant difference in variance between the two groups. Therefore, separate variance estimates were used in calculation of t rate.

Note Regarding Gender Differences and Digit-Span Capacity Effects

Overview

As discussed in the review of the subject selection procedures (see METHODS chapter), sex and treatment group differences with regard to WISC-R forward digit span capacity were discovered prior to the intervention of the strategy training program. Specifically, males (n=29) recalled on average a significantly longer string of digits (M=4.89) than did females (n=18; M=4.16). With regard to treatment groups, the difference between the group with the highest (control group; n=12; M=5.17) and the group with the lowest (long term 2; n=13; M=4.23) average forward digit-span capacity was statistically significant. These differences had the potential for confounding the analysis of primary interest here, the assessment of treatment effects. Any treatment effects that were discovered, for example, might be a reflection of the sex difference in digit-span capacity combined with the distribution of males and females between the subset of groups being compared. Similarly, the advantage enjoyed by the control group with regard to forward digit-span capacity might be responsible for a finding of lack of treatment effects.

The solution to this problem is to explicitly analyze the influence of the sex and group differences. The results are reported below.

Sex Effects

Sex was included with treatment group in two-way analyses of covariance applied to all of the outcome measures discussed above. In no instance was there a significant main effect for sex or a significant treatment by gender interaction. The male and female means for categorization and recall are reported in Table 24 for pictures and Table 25 for prose. The difference between boys and girls with regard to digit-span capacity observed at Time 1 did not have a subsequent effect on the overall performance of the treatment groups. Thus, the presentation of results was simplified throughout this chapter by reporting only one-way (treatment) analyses of covariance.

Digit-span Capacity Effects

The effect of group differences in digit-span capacity was evaluated by including digit-span as a covariate in a series of analyses of covariance identical in all other respects to those reported above. In no case did the statistical controlling of digit-span capacity in this manner have an impact on the analysis of the main effect for treatment group.

The limited impact of digit-span on the categorization and recall of pictures and prose is illustrated by the data in Tables 26 and 27. The percent of variance in categorization and recall that was accounted for by (a) time 1 performance alone, (b) digit-span capacity alone, and (c) time 1 performance plus digit span are reported. In most cases, adding digit-span capacity to the prediction equation results in only a modest increment in the overall power of the model. In those cases where a more substantial increase in the amount of variance accounted for occurs on a percentage basis (for example, the increase from 9% to 12% with regard to the prediction of the number of items recalled from the first pile of pictures at time 3/deck 5), the absolute size of the multiple correlation coefficient remains small. If digit-span capacity represented some mental ability which truly related to use of categorization and/or recall efficiency, one would expect a more consistent if not a greater impact on those measures.

Table 24

Sex Differences in Categorization and Recall of Pictured Items Used as Indicators of Treatment Effect

	Sex			
Task and Stimulus	Male (N)	Female (N)	t (df)	p (two-tailed)
Categorization (ARC scores)				
Subject-sorted Piles				
Deck 1B (Time 1)	.87 (19)	.86 (13)	-0.23 (30)	>.80
Deck 3 (Time 2)	.94 (27)	.93 (16)	-0.21 (41)	>.80
Deck 5 (Time 3)	.89 (19)	.90 (12)	-0.28 (29)	>.75
Fixed-Order Piles				
Deck 4B (Time 2)	.84 (26)	.89 (14)	1.30 (38)	>.20
Deck 6B (Time 3)	.56 (19)	.33 (15)	-1.64 (32)	>.10
Recall (Proportion Items Recalled)[1]				
Subject-sorted Piles				
Deck 1B	.80 (19)	.85 (13)	1.21 (30)	>.20
Deck 3	.77 (27)	.79 (16)	0.78 (41)	>.40
Deck 5	.81 (19)	.79 (13)	0.50 (15.8)	>.60
Fixed-Order Piles				
Deck 4B	.82 (26)	.89 (14)	1.37 (38)	>.17
Deck 6B	.84 (19)	.76 (15)	-0.98 (32)	>.30

[1] Raw proportions are reported in table for ease of interpretation; statistical tests were performed using arcsine transformed proportions.

Table 27

Relationship of Digit-Span Capacity with Categorization and Recall of Pictured Items

Task and Dependent Variable	Percent Variance in Dependent Variable Accounted for by[1]: Baseline Ability Alone	Digit-Span Alone	Baseline Plus Digit	N
Categorization (ARC scores)				
Subject-sorted Piles				
Deck 3 (Time 2)	0.11	4.0	4.1	43
Deck 5 (Time 3)	0.52	2.3	2.4	31
Fixed-Order Piles				
Deck 4B (Time 2)	2.1	0.12	2.5	40
Deck 6B (Time 3)	2.4	5.0	8.5	34
Recall (Proportion Items Recalled)[2]				
Subject-sorted Piles				
Deck 3 (Time 2)	0.28	0.72	1.2	43
Deck 5 (Time 3)	9.1	3.5	12.1	32
Fixed-Order Piles				
Deck 4B	8.0	5.7	13.6	40
Deck 6B	8.6	1.7	9.6	34

1. As represented by the square of the multiple correlation coefficient, multiplied by 100.

2. All calculations computed using arcsine transformation of the raw proportion of items recalled.

Table 25
Sex Differences in Categorization and Recall of Items from Prose Paragraphs Used as Indicators of Treatment Effect

	Sex			
Task and Stimulus	Male (N)	Female (N)	t (df)	p (two-tailed)
Categorization (ARC scores)				
Blocked Paragraphs				
Par 3 (Time 2)	.79 (27)	.80 (16)	0.08 (41)	>.90
Unblocked Paragraphs				
Par 4 (Time 2)	.66 (27)	.69 (16)	0.28 (41)	>.75
Par 5 (Time 3)	.62 (18)	.64 (13)	0.18 (29)	>.85
Recall (Proportion Items Recalled)[1]				
Blocked Paragraphs				
Par 3 (Time 2)	.89 (27)	.91 (17)	0.31 (42)	>.75
Unblocked Paragraphs				
Par 4 (Time 2)	.83 (27)	.88 (16)	0.92 (41)	>.36
Par 5 (Time 3)	.81 (18)	.81 (13)	-0.04 (29)	>.95

1. Raw proportions are reported in table for ease of interpretation; statistical tests were performed using arcsine transformed proportions.

Table 26

Relationship of Digit-Span Capacity with Categorization and Recall of Items from Prose

Task and Dependent Variable	Percent Variance Accounted for by[1] Baseline Ability Alone	Digit-Span Alone	Baseline Plus Digit	N
Recall (Proportion Items Recalled)[2]				
Blocked Paragraphs				
Par 3 (Time 2)	8.2	0.10	8.3	43
Unblocked Paragraphs				
Par 4 (Time 2)	27.2	<0.01	27.2	42
Par 5 (Time 3)	2.8	0.10	2.9	30
Categorization (ARC Scores)				
Blocked Paragraphs				
Par 3 (Time 2)	12.8	3.4	15.2	41
Unblocked Paragraphs				
Par 4 (Time 2)	4.3	5.6	10.8	42
Par 5 (Time 3)	3.2	<0.01	3.3	30

1. As represented by the square of the multiple correlation coefficient, multiplied by 100.

2. All calculations computed using arcsine transformation of the raw proportion of items recalled.

Table 27

Relationship of Digit-Span Capacity with Categorization and Recall of Pictured Items

Task and Dependent Variable	Percent Variance in Dependent Variable Accounted for by[1]: Baseline Ability Alone	Digit-Span Alone	Baseline Plus Digit	N
Categorization (ARC scores)				
Subject-sorted Piles				
Deck 3 (Time 2)	0.11	4.0	4.1	43
Deck 5 (Time 3)	0.52	2.3	2.4	31
Fixed-Order Piles				
Deck 4B (Time 2)	2.1	0.12	2.5	40
Deck 6B (Time 3)	2.4	5.0	8.5	34
Recall (Proportion Items Recalled)[2]				
Subject-sorted Piles				
Deck 3 (Time 2)	0.28	0.72	1.2	43
Deck 5 (Time 3)	9.1	3.5	12.1	32
Fixed-Order Piles				
Deck 4B	8.0	5.7	13.6	40
Deck 6B	8.6	1.7	9.6	34

---

1. As represented by the square of the multiple correlation coefficient, multiplied by 100.

2. All calculations computed using arcsine transformation of the raw proportion of items recalled.

DISCUSSION

The strategy training project attempted to answer certain questions about the relationships between one strategy for information processing--categorization--and the transfer of that strategy to an application in the real world. The transfer task, remembering information from prose, was selected as an example of an everyday situation in which children are called upon to catgorize information or to use information that is already categorized in the text. The outcome variables of interest, then, were: (1) the degree to which categorization was used by the children in the study before and after training, with regard to both a simple categorization task involving picture stimuli, and with respect to information in prose; and (2) the degree to which recall of information was improved as a result of training.

The specific hypotheses tested in the present research were as follows:

1. A measure of the use of categorization in recalled picture stimuli (ARC scores) would be: (a) higher for treatment than for non-treatment controls, and (b) higher for those receiving more training, i.e. higher for those children in the long-term training conditions than the short term condition.

2. A measure of the use of categorization in recalled items from prose stimuli (ARC scores) would be: (a) higher for treatment than for non-treatment controls; (b) higher for long-term treatment groups than for the short-term treatment group.

3. Free recall scores calculated as the proportion of items recalled from sets of picture stimuli would be: (a) higher for treatment than for non-treatment controls, and (b) higher for children in long-term training conditions than for children in the short-term training condition.

4. Free recall scores, calculated as the proportion of items recalled from prose passages, would be: (a) higher for treatment than for non-treatment children, and (b) higher for children in long-term training conditions than for children in the short-term training condition.

These hypotheses regarding the effects of training in the use of a categorization strategy were, in general, not borne out. Specifically, there were no differences in the degree of categorization evident in the recall protocols of trained and control groups with respect to picture stimuli. There were also no significant differences between the scores of long-term and short-term training conditions with respect to this measure (ARC). Similarly, there were no significant differences between the mean ARC scores of the treatment and control groups with respect to prose stimuli, indicating similar degrees of use of the categorization strategy in both groups. This result was consistent across passages in which the items in a given category were blocked, and prose stimuli in which the category members were distributed thoughout the passage (that is, unblocked).

That is not to say, however, that training did not have an effect. Inspection of the means of all four groups reveals that ARC scores for picture stimuli and prose stimuli increased from the pre- to the post-treatment assessment. The crucial point is that these increases were evident in all four groups, including the no-treatment controls who had been exposed to a brief demonstration of categorization at the conclusion of the pretest session (see METHODS discussion above).

With respect to recall efficiency, the hypotheses about the effects of this strategy training curriculum were also not borne out. Analyses of results of pre- and post-treatment scores for recall efficiency for both subject-sorted picture stimuli and fixed-order picture stimuli revealed no significant differences between treatment and control groups. The comparison of long-term and short-term training conditions revealed that for subject-sorted picture stimuli, there were significant differences between training conditions: subjects who received four weeks of training recalled more of the picture stimuli than did subjects who received two weeks of training. However, for the fixed-order picture stimuli, subjects who received two weeks of training performed significantly better than did subjects who received four weeks of training. Although there were significant differences in short-term memory capacity between these groups at the pre-test assessment, analysis of covariance (adjusting for these pre-training differences) demonstrated that the short-term training group recalled more items from fixed-order picture stimuli than the long-term training groups immediately after training.

Finally, the hypotheses concerning recall efficiency for prose stimuli were not confirmed. Comparisons between treatment and controls showed no significant effects of training, and comparisons between short-term and long-term training conditions demonstrated that extent of training had no significant impact on recall of items from either blocked or unblocked prose stimuli.

While the results of these analyses are relatively unequivocal with respect to the effects of this curriculum, we do not conclude that training in cognitive strategies is ineffective--either in its own right, or as an integral part of a reading comprehension curriculum. Rather, there were a number of factors in operation during the data collection which certainly influenced the results of the study. These factors may roughly be divided into practical and design-related concerns. On the practical side, there were some factors which affected data collection in this study which were unique to the school climate and timing of the data collection which, although interesting from a methodological point of view, do not shed light upon children's abilities to learn and apply cognitive stragegies for reading comprehension. There are other factors involved in the specific design of the experimental task and conditions of assessment that do shed light upon children's acquisition and use of cognitive strategies. All these factors are discussed below, beginning with the primarily practical issues.

Practical Issues

School Climate

As discussed above, the school in which this research took place was a large urban elementary school in a low-to-middle income neighborhood of a working-class city near Boston. The research team was unprepared for the level of unpredictability at all levels of this particular school. Some of this unpredictability had to do with school politics. Because of massive budget cuts, school personnel were being reassigned throughout the year, and morale at the level of the school administration and at the level of the classroom teacher was very low. As a result of unforeseen delays in dealing with the school administration, the study was conducted later in the school term than we would have liked, with final data collectiooccuring during the last weeks of school.

There was also a great deal of unpredictability with respect to the children and their performance in school. Truancy was a constant problem at this school, so data collection was made more difficult by the subjects' frequent absences. Also, as noted above, there were children who stopped coming to school entirely when they discovered toward the end of the school year that they were not going to be promoted to the next grade level. Although this amount of unpredictability is probably the norm in this school, and perhaps other urban schools like it, it was a factor that affected our data collection for which we were unprepared. As a consequence, some of the cells in the analyses reported above have less-than-optimal numbers of subjects, and some of the results may be concomitantly affected by "outlier" performances among the children who remained in our pool throughout the entire study.

Peer "contamination" effects. One consequence of the school climate noted above was a lack of one-to-one contact between teachers and children: because of large class sizes, individual tutoring was an infrequent occurrence. The tutorial sessions with trainers from the Strategy Training Project became, almost witout exception, highly valued meetings. Children got out of their regular classes and were exposed to something new and different under the guidance of a trainer who was warm, sympathetic, and who gave the child a great deal of support for whatever achievement he or she managed in the training sessions. Children who were selected as no-treatment controls were in the same classes as the children who were in training conditions, and there appeared to be strong feelings about who was allowed to go to tutorial sessions and who was not. Project personnel, both trainers and testers, heard anecdotes from study participants about children giving or receiving training to each other to "share the wealth" of what was going on in the tutorial sessions. This phenomenon is not unknown in educational research: for example, Rist (1970) reported a similar phenomenon in low-SES schools in which students thought to have promise were found to be teaching lessons to their less "promising" peers after school.

In future studies, experimental design can be refined to control for the possbility of peer teaching effects. These design changes include: (a) including a no-treatment group at a comparable but different school to systematically investigate the presence of peer-teaching effects; or (b) including an attention-placebo control group at the same school, so that control children receive some of the perceived good effects of training sessions and the demand, among children, for peer training is modified.

Design-Related Issues

Category cueing and the effects of pre-test procedures.

The most powerful explanation for the lack of significant findings in the present research has to do with the power of the categorization demonstration that concluded the pre-test assessment session for all subjects. In initial phases of the pre-test, children were given picture stimuli in a movable array (one picture per card) and told that they would be asked to remember the pictures later. They were told to place the pictures in an order such that the order they selected would help them to remember.*

At the end of the pre-test session, after the assessment of categorization and recall in fixed-order piles and in blocked and unblocked prose stimuli, each subject was asked to return to their original deck of pictures. The suggestion was given to put the pictures in groups of "things that go together" as a way to aid in memorization. The tester used a new deck of 4 picture cards, consisting of two items in each of two categories, to demonstrate the idea of categorization. The subjects' subsequent free recall performance, as measured by the ARC score, showed a dramatic increase in extent of categorization across all experimental groups (i.e., treatment groups and control groups). Table 6 shows the effects of this instruction by comparing performance in the use of categorization before and after the "category cue" (demonstration). ARC scores before the category cue ranged from .34 for subjects in one long-term group to .78 for subjects in the short-term group, with a mean ARC score for all groups of .54. After the category cue, ARC scores ranged from .79 for the control group to .95 for the short-term treatment group. Thus, the subjects were using the strategy of categorization almost perfectly at the end of the pre-test procedure. As noted above, paired T-tests for these pre-post differences remained significant over the course of the study. With the scores of all subjects, including the no-treatment controls, pooled, there were significant differences between post-test 1 (two weeks aftter pre-test) and pre-test, as well as significant differences between post-test 2 (four weeks after pre-test) and pre-test on ARC scores for both subject-sorted and fixed-order piles (Table 7).

---

*This instruction, while slightly complicated and inelegant, reflects the difficulty of suggesting to subjects that they organize the pictures in some ways without suggesting "grouping" or any other cue to the use of categorization. See, for example, Borkowski (1983).

Clearly, the category cue was sufficiently powerful to wash out any differences between groups that been associated with training. In a sense, the "training" in this study appears to have been largely accomplished with the category cue given in the pre-test session. In all cases, initial training led to gains over and above the initial leap in performance following the category cue. For example, after two weeks of training, all subjects improved in their ability to use categorization to recall items from blocked and unblocked prose (Table 9). But the difference between control subjects and subjects in treatment groups was not significant.*

These data also consistently indicate decrements in performance in the absence of continued training. For both the control group and the short-term training group, mean scores decrease for categorization in recall and recall efficiency at Time 3 vis a vis Time 1 (i.e., immediately after the four-week training sessions but two weeks after the end of the short-term condition and four weeks after the initial pre-test session). It thus appears that training does have an effect, but the effect is not sufficiently powerful to elevate the performance of those who were trained a statistically significant amount above those who were only exposed to the

---

*Note that this contrast assesses subjects before the "category cue" on pre-test, because the pre-test on prose passages came before the demonstration of categorization. Thus Time 2 (two-week post-test) is the control subjects' first assessment of prose passages following the "category cue."

"category cue." As a final example, note that in one experimental group (the short-term training group), the ARC score for categorization in recall was .95 for the task immediately following the categorization demonstration at the end of the pre-test session. Since a perfect ARC score, reflecting perfect categorization, is 1.0, it is difficult to imagine how training could substantially improve performance on this variable.

Memory Load

A related issue is the issue of load with respect to the subjects' short-term memory system. Clearly categorization, or any memory strategy, is only required as load increases; if load is sufficiently small, short-term or "brute force" memory can be called upon for accurate recall. The implication is that if memory tasks are sufficiently easy, training in categorization will not affect recall because the strategy is not needed to perform adequately.

It appears that some of the stimulus arrays used in the present study were not sufficiently taxing to induce memory strategy use. For example, blocked prose passages, in addition to being easy because the items were already arranged in categories, consisted of nine items in three categories. This is only slightly above the "magic number" seven (Miller, 1956) that most adult subjects can remember in short-term memory. Inspection of the results in this study indicate that the proportion of items recalled from blocked prose passages ranged from 84 to 87% correct before training of any kind, including the category cue. Although long-term subjects improved, from 85 to 96% recall, while control subjects remained at their pre-test level of 87%, this degree of variability is not sufficient to demonstrate significant effects of training. ARC scores for categorization in recall from blocked passages of prose ranged from .41 to .67. This variability in the use of categorization, when contrasted with the uniformly high recall performance just noted, suggests that the strategy of categorization was not required to achieve a high recall score.

A formal measure of the relationship between use of the categorization strategy and recall efficiency is available in the correlation between these variables on any given task. The Pearson product-moment correlation between ARC score and amount recalled for blocked prose passages ranged from .06 for the short-term training condition to .45 for one of the long-term conditions. None of the correlations achieved statistical significance.

Taken together, the effect of the "category cue" on all subjects and the effect of memory load conditions, which may not have been sufficient to require strategy use, may have combined to obscure real differences between experimental groups in their response to training in categorization. In the future, we intend to take these factors into account by including measures of response to treatment that are more in keeping with the main focus of the training sessions themselves. In fact, because of the immediate response to the category cue, very little time was spent during the subsequent training sessions on the idea of using categories for remembering, especially in the domain of simple stimulus arrays (with a few exceptions; see below). Measures of more complex prose stimuli with larger numbers of items may demonstrate the effects of this strategy training curriculum more effectively by preventing subjects from scoring well due to simple short-term memory. It is in this area that future assessments need to be designed.

Group and Individual Differences

There were seven subjects in the three treatment groups who required extensive training in categorization before they were able to meet a pre-established performance criterion and thereby advance to the next stage in the curriculum, the application of categorization to prose material.* Two of the "uninsightful" children were in the short-term training condition; the remaining five were all in the same long-term training condition. The differences in performance between these seven children and the remaining 28 treatment subjects were striking. Although there were no significant differences in IQ or WISC digit span, there were significant differences between "insightful" and "uninsightful" children on a number of outcome variables (see Tables 20 and 21). For example, in blocked prose passages, pre-test scores for categorization (ARC score) and recall (proportion recalled) were .57 and 87% for "insightful" chilren and .04 and 71% for "uninsightful" children. At the first post-test, these scores were .83 and 90% for "insightful" children and .56 and 76% for "uninsightful" children. All of these results represent statistically significant differences (at the .05 level, one-tailed; t-test for difference between independent group means).

These results indicate striking individual differences in children's response to the intervention, differences which have implications for the design of future research efforts and for the application of cognitive strategy in general. There appear to be differences among children in their ability to "pick up" or use cognitive strategies even after they are provided. Such differences in ability to learn cognitive strategies (specifically, differences in rate of acquisition of cognitive strategies) were independent of IQ as measured in this study. The methodological implication of this finding is that the experimental groups were not truly matched in the present research. Although the groups were matched on the IQ variable, and treatment conditions assigned to groups after this matching process, it appears that a better test of the strategy training curriculum might have been achieved if subjects were matched on their ability to profit from cognitive strategy training. In the present study the distribution of "uninsightful" children was uncontrolled. As noted above, it is impossible to know how many "uninsightful" children were contained in the control group. The distribution of such children across the other three treatment groups may, however, have contributed to the anomalous results with respect to these groups' poor response to training.

---

*Since the difficulty that these children demonstrated regarding the acquisition of the categorization strategy only became evident in the course of actual training, we cannot know how many subjects in the no-treatment control group would also fall into this group.

For future research efforts, the individual differences reported here have implications for the assessment of children with learning difficulties. The differences reported here are, of course, similar to differences in learning potential (LP) which have been part of the research literature in mental assssment for some time (Budoff, 1969). Assessment strategies that do not include some dynamic component will clearly not give an accurate picture of a child's mental abilities. Research in cognitive strategy training is proceeding on a number of fronts to attempt to investigate the relationships betwen cognitive strategy use and common problems of information processing, like reading comprehension. The individual differences reported here suggest that relationships between cognitive strategy acquisition, strategy use, and problems in information processing are also important. In sum, individual differences in strategy acquisition may be an important dimension in the understanding of learning disabilities, and certainly is an important variable to consider in the design and implementation of interventions intended to further illuminate theoretical or applied issues in cognitive strategy training.

It is also important to note that the existence and performance of the "uninsightful" children have practical implications for remedial reading interventions in addition to those just discussed with regard to research design.

Specifically, even though they did not improve their categorization and recall performance sufficiently to close the significant gap between themselves and the insightful poor readers in the study, it should nevertheless not be overlooked that the uninsightful poor readers did improve as a result of their exposure to strategy training. For example, the extent of categorization evident in fixed-order picture piles improved from an ARC score of .20 at Time 1 to .60 at Time 3. With regard to prose, categorization jumped from an ARC score of .04 at Time 1 to .56 at Time 2 (for blocked paragraphs; with regard to unblocked prose, performance was esssentially unchanged over time at an ARC score level of approximately .35; see Table 22). The point to be made here is that a curriculum such as that used in the strategy training project, which involved simpler "back-up" tasks at each stage in the development and extension of the principle of categorization, was able to improve the performance of even those children who demonstrated the least ability to grasp and apply the cognitive strategy at the outset.

Thus, while the results of the study reported here were disappointing in some respects, there were clear indications for future research to continue investigations of the application of cognitive strategy training. In future research, changes should be made in assessment procedures, to fill out the picture of the abilities and disabilities of experimental subjects, with an eye toward better matching of experimental groups and better design of strategy training curricula. Changes in assessment procedures should also take into account the power of strategy cueing, and the necessity of increasing memory load in order to fully activate strategy use. Changes in the design with respect to better control conditions should allow real differences between trained and untrained groups to emerge. Finally, a full investigation of the conditions necessary for maintenance and generalization of trained strategies, not possible as part of this research due to practical limitations of time and resources, will be an integral part of our next research effort, in order to make our research efforts useful to handicapped learners. By investigating the conditions necessary for classroom application of cognitive strategies, we hope to continue our efforts to bridge the gap between basic and applied research and to help handicapped learners benefit from current research efforts in this field.

REFERENCES

Atkinson, R.C., & Schriffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), The psychology of learning and motivation (Vol. 2). New York: Academic Press.

Belmont, J.M., & Borkowski, J.G. (1978, March). Instructing retarded children on two examples of a memory method (vs. only one) improves maintenance and generalization of that method. Paper presented at the Eleventh Annual Gatlinburg Conference on Research in Mental Retardation, Gatlinburg, Tenn.

Belmont, J.M., & Butterfield, E.C. (1977). The instructional approach to developmental cognitive research. In R.V. Kail, Jr. & J.W. Hagen (Eds.), Perspectives on the development of memory and cognition. Hillsdale, NJ: Lawrenece Erlbaum Associates.

Bilsky, L.H., & Evans, R.A. (1979). Clustering and categorical list retention in the mentally retarded. In N.R. Ellis (Ed.), Handbook of mental deficiency, psychological theory and research (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bock, R.D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw-Hill.

Borkowski, J.G., & Cavanaugh, J.C. (1979). Maintenance and generalization of skills and strategies by the retarded. In N.R. Ellis (Ed.), Handbook of mental deficiency: Psychological theory and research. Hillsdale, NJ: Lawrence Erlbaum Associates.

Borkowski, J.G., Cavanaugh, J.C., & Reichart, G.J. (1978). Maintenance of children's rehearsal strategies: Effects of amount of training and strategy form. Journal of Experimental Child Psychology, 26, 288-298.

Bousfield, W.A. (1953). The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 49, 229-240.

Bower, R.H., Lesgold, A.M., & Tieman, D. (1969). Grouping operations in free recall. Journal of Verbal Learning and Verbal Behavior, 8, 481-493.

Brown, A.L. (1978). Knowing when, where, and how to remember: A problem in metacognition. In R. Glaser (Ed.), Advances in Instructional Psychology (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A.L., & Campione, J.C. (1977). Training strategic study time apportionment in educable retarded children. Intelligence, 1, 94-107.

Brown, A.L., Campione, J.C., Bray, N.W., & Wilcox, B.L. (1973). Keeping track of changing variables: Effects of rehearsal training and rehearsal prevention in normal and retarded adolescents. Journal of Experimental Psychology, 101, 123-131.

Brown, A.L., Campione, J.C., & Day, J.D. (1981). Learning to learn: On training students to learn from texts. Educational Researcher, 10(2), 14-21.

Brown, A.L., Campione, J.C., & Murphy, M.D. (1974). Keeping track of changing variables: Long-term retention of a trained rehearsal strategy by retarded adolescents. American Journal of Mental Deficiency, 78, 446-453.

Brown, A.L., & DeLoache, J. (1978). Skills, plans and self-regulation. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ: Lawrence Erlbaum Associates.

Budoff, M. (1969). Learning potential: A supplementary procedure for assessing the ability to reason. Seminars in Psychiatry, 1, 278-290.

Butterfield, E.C., Wambold, C., & Belmont, J.M. (1973). On the theory and practice of improving short-term memory. American Journal of Mental Deficiency, 77, 654-669.

Campbell, D.T., & Stanley, J.C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campione, J.C., & Brown, A.L. (1977). Memory and metamemory development in educable retarded children. In R.V. Kail, Jr. & J.W. Hagen (Eds.), Perspectives on the development of memory and cognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Case, R. (1978). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R.S. Singer (Ed.), Children's thinking: What develops? Hillsdale, NJ: Lawrence Erlbaum Associates.

Chall, J. (1979). The great debate: Ten years later, with a modest proposal for reading stages. In L.B. Resnick & P.A. Weaver (Eds.), Theory and practice in early reading. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M.T.H. (1978). Knowledge structures and memory development. In R.S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dalrymple-Aleord, E.C. (1970). The measurement of clustering in free recall. Psychological Bulletin, 1, 32-34.

Ellis, N.R. (1970). Memory processes in retardates and normals. In N.R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 4). New York: Academic Press.

Fisher, M.A., & Zeaman, D. (1973). An attention-retention theory of retarded discrimination learning. In N.R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 6). New York: Academic Press.

Flavell, J. (1970). Developmental studies of mediated memory. In H. Reese & L. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 5). New York: Academic Press.

Flavell, J.H., & Wellman, H.M. (1977). Metamemory. In R.V. Kail & J.W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhn, D. (1974). Inducing development experimentally: Comments on a research paradigm. *Developmental Psychology*, *10*, 592-600.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *32*, 301-345.

Rist, R.C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, *40*(3), 411-451.

Robinson, F.P. (1941). *Diagnostic and remedial techniques for effective study*. New York: Harper & Brothers.

Roenker, D., Thompson, C.P., & Brown, S.C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45-48.

Turnure, J.E., Buium, N., & Thurlow, M.L. (1976). The effectiveness of interrogatives for promoting verbal elaboration productivity in young children. *Child Development*, *47*:851-855.

Turnure, J.E., & Thurlow, M.L. (1973). Verbal elaboration and the promotion of transfer of training in educable mentally retarded children. *Journal of Experimental Child Psychology*, *15*, 137-148.

Worden, P.E. (1983). Memory strategy instruction with the learning disabled. In M. Pressley and J.R. Levin (Eds.), *Cognitive strategy research: Psychological foundations*. New York: Springer-Verlag.

Appendix A

Letter to Teachers Describing the Target Child of the Strategy Training Project

RESEARCH INSTITUTE FOR EDUCATIONAL PROBLEMS, INC.
29 WARE STREET ■ CAMBRIDGE, MASSACHUSETTS 02138 ■ TELEPHONE (617) 868-0360-2

1/12/83

STRATEGY TRAINING PROJECT: THE TARGET CHILD

The RIEP Strategy Training Project is seeking to apply cognitive strategies for remembering information to the problems of poor readers, in an attempt to increase reading comprehension. Therefore, we are seeking children who can read, but do not seem to be able to remember or comprehend what they have read. The following guidelines will help us to identify which children in your class might be appropriate for the project:

1. Children who are poor readers primarily with weaknesses in organization, comprehension and retention skills. As a rough guideline, we are defining "poor readers" as those students who are reading at least two grade levels below their current grade.

2. Children who are decoding at at least a third grade level. Because the project is aimed at comprehension skills, we are looking for children for whom decoding is not the major reading problem.

The children you suggest can be receiving special education services for reading problems or other learning disabilities. Although some children may turn out to be ineligible because of time requirements of resource rooms, etc., we are, for now, just looking for a rough estimate of all the children that might be involved in the project.

On the attached sheet, please write the names of students in your classes that fit the above description. Please indicate, for each child, what special education services they now receive, if any. If you have any questions or comments, write them on the attached sheet or call one of us at 868-0360. Thank you for your time.

David Anderegg
Cathy Cuneo
RIEP Strategy Training Project

Appendix B

Strategy Training Curriculum Manual:

Pilot Study Version

Susan Kelley

Research Institute for Educational Problems

STRATEGY TRAINING OUTLINE

1.0 Attitude of the Trainers

2.0 The Initial Training Sequence

    2.1 Overview of the Initial Training Sequence

    2.2 Picture Task I (PT 1)

        2.2.1 Link to the pre-test

        2.2.2. First trial at PT 1

        2.2.3 Performance criterion for PT 1

        2.2.4 Analysis of student processing

        2.2.5 Shaping of student processing to fit the stragegy

    2.3 Picture Task II (PT II)

    2.4 The Sequence of Back-up Tasks

        2.4.1 Back-up task 0 (BT-0)

        2.4.2 Back-up task 1 (BT-1)

        2.4.3 Back-up task 2 (BT-2)

    2.5 The Initial Sequence of Prose Tasks

        2.5.1 Bridge from the picture tasks

        2.5.2 Preparatory steps for prose

        2.5.3 Prose task I (Pr I)

        2.5.4 Prose task II (Pr II)

        2.5.5 Prose task III (Pr III)

        2.5.6 Prose task IV (Pr IV)

    2.6 Making a List of Rules for Grouping

        2.6.1 "Rules for Making Groups"

    2.7 Conclusion of the Initial Training Sequence

3.0 The Continued Training Sequence

    3.1 The long-term condition

    3.2 The supportive curriculum condition

    3.3 Overview of the continued training sequence

    3.4 The seven concepts taught in the continued training sequence

1.0 Attitude of the Trainers

Trainers provided an interpersonal context in which the students' learning was well supported. They made every effort to communicate (1) respect for the student and for the student's performance, (2) curiosity about the information processing the student used, (3) belief in the notion that each student's mental process (a) was unique and (b) had untapped potential, and (4) the expectation of positive outcome in student performance. Throughout the training, failure to meet grouping or memory criteria (see 2.2.3) was defined as feedback, that is, information helpful for improving student performance by providing information about the student's mental processing.

2.0 The Phase I Training Sequence

2.1 Overview

Training began with the task of grouping and remembering 20 items shown on stimulus cards. The items belonged to five mutually exclusive categories. This task, called Picture Task I, was identical to that attempted in the pre-test, except that it was slightly more difficult, requiring the learning of 20, rather than 15, items.

This initial task lay on a continuum of tasks, some more and some less difficult than the original task. For each student, the appropriate beginning task was found through trial and error. The appropriate beginning task was the task at which the student first met the criterion for successsful performance for that task. Beginning at Picture Task I, it was possible to move in the direction of easier tasks--called back-up tasks--as well as in the direction of more difficult tasks. (When competence was not initially achieved, then the trainer backs up to a less complex or demanding task.) After the appropiate beginning task was found, each student proceeded in the direction of tasks of increasing difficulty. A flow chart (see illustration II) showing the progression of training tasks illustrates this 2-way route. It also shows the optional tasks (Prose tasks IH, IIH, IIIH, and IV) provided for those students who moved through the tasks more quickly. These optional tasks were used whenever possible, providing horizontal elaboration of the curriculum for almost all the students in the study.

The training sequence provides an orderly progression through tasks of increasing difficulty, yet is flexible enough to allow for variation among students.

2.2 Picture Task I

As was mentioned above, all students began training by attempting to remember the 20 items pictured on the stimulus cards of Picture Set I. The trainers did not offer any aid to students in this first attempt, but closely observed them, watching for behavioral cues to mental process. After the 20 cards were laid out in random order, the task was introduced in the following way:

2.2.1 Link to the Pre-test

The trainers said:

> "Remember what you did on the pre-test?
>
> Well, you have the same job to do with these cards.
>
> You put them in a pile, and arrange them so it's easy for you to remember them all. You can put them in any order you want.
>
> Only this time, it is not a test.
>
> What we are going to do is learn some new ways of doing this job, so you can do it better and easier.

But first, I'd like to see how yu do this right now."

2.2.2 First Trial at Picture Task I

A variable time period was allowed for each student to self-pace organizing and memorizing the cards. The trainers used their judgment in cutting short excessively long preparation periods. The trainers found overstudying and reluctance to attempt recall were common among the students in the study. Therefore, after what seemed to be a reasonable time, the trainers told reluctant students, "Well, let's try it [recall] and see what happens."

During student recall, trainers recorded items as they were uttered, numbering each response. The spatial arrangement of the recorded items, however, was kept to clearly separated group lists. (See illustration.) This recorded list was kept out of the student's line of sight.

2.2.3 Performance Criterion for Picture Task I

The performance criterion for Picture Task I was the recall of 3 out of 4 items as a group for 3 of the categories. Items were considered recalled as a group if they were spoken in sequence or if a student, recalling a single item, said: "Oh, this goes with [these others]." or "Oh, I remember another [category item]. It's X."

If the student failed to meet the performance criterion for Picture Task I, the trainers said:

> "That was a good try.
> Now let's see what you can do with these cards."

In this way the student was routed onto the backup sequence (see 2.4).

If the student met the criterion, the training continued without backtracking (see 2.2.4).

2.2.4 Analysis of Student Process

After the student had met the criterion for Picture Task I, the trainers congratulated him for doing so well "without having any of our lessons yet," and then said:

> "I'd like to know how you did that?
> What was in your mind when you were doing that?
> Do you know how you did that?"

The trainers encouraged students to use the stimulus cards to illustrate their processing descriptions.

Trainers kept in mind that the students' descriptions of their internal process may not accurately reflect the characteristics of that process. These descriptions were considered only partial information, and were combined with careful observation of student behavior by the trainers, who tailored individualized approaches for teaching the components of the strategy.

The cognitive style of the individual student was the guide for designing specific interventions. The trainers oriented their teaching to visual, auditory, or kinesthetic emphases. The repertoire of interventions was varied. Specific interventions were not as important as the match of the intervention mode with the student's cognitive style, allowing for easier assimilation by the student. Thus, for example, in memorizing items within a category, one student may have found it more effective to visualize the items or a list of the items; another student to repeat the names of the items in each group several times over; and a third to manipulate the cards while sub-vocally mouthing the names of the items.

2.2.5 Shaping of the Student Process to Fit the Strategy

The verbalization of process by the students served to alert the student that he was using a process to which he had conscious access. It was the beginning of the conscious acquisition of the strategy. The teaching of the strategy then began; the focus of training sessions was the shaping of mental process. This soon became a matter of mutual interest and effort for both trainer and student.

Steps in this shaping procedure will be described in some detail below; much of the learning of the strategy took place within the work on the Picture Task; therefore this task is considered in detail. It represents the foundation of the training. Further training sessions refined the strategy taught in the beginning of training, and expanded it to prose materials.

Remember that at this point in the training all students had met the criterion for Picture Task I and therefore were using some acceptable form of grouping procedure. In addition, students had attempted to verbalize this procedure in response to questioning from the trainers. Trainers built on this grouping behavior, meeting each student at his particular level in the development of the categorizing concept.

In introducing the shaping process, trainers used analogies which supported the notions of (a) successful intervention and (b) control over process; for example, the students' processes were likened to

> "A car which is running all right. It will get you where you want to go, but if you take it to the garage and get it tuned up--get it a really good tune-up--then it's really going to purr along. That fine-tuned engine will run so smooth. And it will be more fun and easier for you to drive better."

Trainers presented the third step in the shaping process with some drama, underscoring its importance. The trainers found this the most complex segment in the strategy to teach, because it involved careful negotiation of several stages. It was of particular importance to adjust the pace of the teaching to the student's rate of increasing comprehension during this step. This step involved (a) eliciting group names from the student, and (b) substituting a list of five group names for the list of 20 items as the primary focus of memorization.

The *first step* in the shaping of student strategies, then, was the *firming-up of the concept of groups*, using the specific images of PT-1. *Work in this phase was limited to the student's arrangement of stimulus cards in his pile, and did not extend to the recall list*. Trainers asked students why certain cards had been placed in proximity. "Why did you put this card on top of this one?" Any differences in the order of student arrangement and the ideal mutually exclusive order of the 20 cards were discussed. Trainers catalyzed this discussion by asking leading questions like: "Do you think this card goes with any of the others?" In the case of a solid difference of opinion in the placement of items, the student's rationale was accepted and his placement retained.

The *second step* in shaping was *creating distinct and firm boundaries between the groups*. The large pile of cards was physically separated into five group piles, each placed a good distance from the other four piles. The five piles were spread across the able in front of the student. The self-contained character of each group pile was underscored, using a technique which matched each student's cognitive mode. For kinesthetic learners each pile was lifted, cards manipulated, and returned with a distinct "thump" to the table; for auditory learners, the trainers said, "Now the a, b, c, and d [items making up one pile]"--trainers made an emphatic pause--"and the e, f, g, and e [items of the other pile] are good and separate"; for visual learners, the trainers said, "Now you can easily see how each pile is separate from the others."

To elicit group names, trainers asked, referring to the items of one group, "What is one word you could use to describe all these things?"; or, "If these things were all piled up in the sink, and I said to you 'Will you please wash those dirty _______, what would I call them?"; or, "If I were to go to a store to buy these things, what kind of a store would I go to?" and so forth. The hints in the trainer's leading questions were increased as needed until a group label was produced by the student.

After the student had produced a group label, the trainer began a short discussion with the student as to the "goodness" of that label. This "discussion" was held whether or not the trainer considered the label in question a "good" label. In this way the trainer modelled a "discussion" the student could later have within himself to ascertain the appropriateness of future group labels.

Such a discussion might go like this:

> *So you put the slide, the wing, the seesaw, and the sand box into the same group.*
>
> Right.
>
> *And I agree that they do go together. - And you decided to call them "toys" for a group name.*
>
> Yup.
>
> *Well...do you think that will work for a group name?*
>
> What do you mean? They are all toys, aren't they?
>
> *Yes, they are, I suppose. But do you want electric trains or dolls or Legos in that group?*
>
> No.
>
> *Why not?*
>
> Well, because Legos and stuff like that is different.
>
> *But they're all toys.*
>
> Yeah... This is hard. I don't know what to call them.
>
> *Would you let a jungle gym into your group?*

Yes, I guess I would.

Why?

Because that's something you play with outside too. You know, like at the park or a playground or something.

So you want to let things you play on at the park or playground into your group, but you don't want to let in toys like dolls and Legos?

Right.

So what can you call this group? Think of a name that will let in all the things you think should get in, and will keep out all the things you think should stay out of this group.

I'll call it "Things you find to play on in the park."

Hmm... That keeps out the electric trains, and it tells what the things in the group are also. Okay, that's a good group name. Do you want to make it shorter?

Well, I could call it "Playground."

After five group labels had been elicited and agreed upon, the trainers introduced phase (b) dramatically:

> "Now I'm going to tell you the secret of doing this."
>
> "Now I am going to teach you the trick to this."
>
> "Now I will show you what I have to teach you which will make it easier for you to do better."

Picking up one group pile and placing it in the center of the table, the trainers announced:

> "I will now show you how to turn four things into one thing."

[SEE ILLUSTRATION IV ABOUT HERE]

The trainers then held up four fingers, counting them off, saying, "One, two, three, four"; and then drew the four fingers down behind the palm of the other hand, changing finger position while out of the student's view, and revealing only one finger as the first hand reappeared. This action and the words "Four - into - One!" were managed in the manner of a magic show trick. (See illustration.) This "magic trick" was then repeated, using the cards of the group pile from the center of the table instead of fingers: "Four things--a, b, c, d [trainer lists item names]--become one thing [group label]." (See illustration.) At this point, the students were told simply and directly that they need not remember all 20 cards, but only the five group names.

The fourth step in the shaping process taught the two-part memorization component of the routine. The five group piles of cards were well separated on the table surface. Trainers directed the students to think about only one of these piles at a time. The trainers said:

> "Choose any one of these piles.
>
> It doesn't matter which one, any one you want."

After the student chose a pile, the trainers moved the other four piles to one side. The trainers said:

> "Now think about just this group you have picked. What are the four things in this group?"
>
> [Student's answer]
>
> "What is the name for this group?"
>
> [Student's answer]
>
> "Now put the cards face down on the table and tell me the names of the things in this group."

With this amount of attention to detail, trainers "walked" the students through the memorization of each group category. Great emphasis was placed on the necessity to keep each group separate from the others. Metaphors were used to facilitate the separate storage of the groups. The human brain, for example, was pictured as having a room full of shelves or cabinets:

> "Now you know all four things in the [clothes] group. Put those things on a shelf up in your brain. You have lots of extra room there, we don't use a whole big space in our brains, nature gives us all a lot of extra brain-space. So

> put all those [clothes] on an empty shelf, and close the door of that shelf. You can write the group name, ["Clothes"], right on the door of that shelf."
>
> [Pause and wait for imaging by student]
>
> "Okay, did you do that?
> Now, you don't need to think about the [clothes] group any more for now. Just leave it up there. It's not going to walk off the shelf!"
>
> [Trainer shakes head a bit, or wipes brow as if to brush off cobwebs.]
>
> "Now we'll do the same thing with the next group."
>
> [Student chooses another group pile and repeats memorization process.]

This process was repeated until all five groups had been learned. The trainers then remarked that the student "now has five labelled doors in your brain," and inquired what the words on those doors were. In response, the student repeated the list of group labels. If a student had trouble remembering the labels, care was taken to work with the student to find a memnonic which was useful in facilitating recall of the list of group names.

The fifth step in the shaping process was recall of the 20 item list. Trainers once again wrote down the items as the student recalled them. This time, however, the trainers placed the paper on which the list was being recorded in view of the student. Enclosing each group within a rectangle drawn on the paper, the trainers paced the student's recall, using prompts and suggestions when necessary to stimulate the recall process. For example, if a student had recalled three items from a group and could not recall the fourth item right away, the trainer would enclose this group, draw a line for the fourth item to be written in later, and say,

> "That's just fine.
> You can go back and get this fourth thing later.
> It's still up there on the shelf.
> Let's go on to another group now."

For each group, the student was asked to supply the group label if he did not do this spontaneously. In general, the recall process was shaped to a two-part routine, in which the recall of the list of group labels was one part, and the filling in of the items of each group was the second part. The recall of the group labels was presented as the more important of the two parts.

Failure to produce items within a group was treated as "a mere detail" which could be "cleared up" by any of a variety to techniques. The trainer and student worked together on these. Students learned to try these "fill in the blank" exercises one after the other until "the one which will work" was finally found. Some of these recall techniques were: repeating the list of items aloud ("this, this, this, and ________"); imagining the cards laid out in front of the student; imagining dealing the cards out one after the other, and so forth. The recall-aid technique in most cases was matched to the original memorization technique. (See step two above.) One recall technique was taught to all students: that was the calling to mind and checking for a "match" items which could logically be considered part of the group in question, based on its category label.

If all of these recall techniques failed to produce a "lost" item--which was called a "misplaced item" by the trainers--the failure was attributed to an oversight in the original memorization of that item. The student was directed to memorize the items over again if the number of "lost" items was large (more than five or six over all categories). Trainers used their judgment in this matter, taking into account student patience, student effort, etc. The student was assured that efffective memorization would bring better recall. The student's attention was directed to the successfully recalled items. Trainers compared the student's first attempt to recall the list of 20 items and

BEST COPY AVAILABLE

this, the second attempt, pointing out that in the second attempt:

(1) more items were recalled;

(2) it was easier to remember the items;

(3) the items were all in order in the student's mind

In the first attempt to recall the items of PT I, some students produced a recall order which jumped from category to category. In this case, the trainers showed the student the two sheets on which the recall had been recorded--one sheet from the recall before strategy-shaping took place, and one sheet from the post-shaping recall. In each case the student's responses had been numbered in order of recall, and in each case the items had been arranged in separate groups. As if playing "follow the dots," the trainers connected the numbered responses in order, producing a meandering, discontinuous line for the first attempt, and an orderly progression for the second.

[SEE ILLUSTRATION III ABOUT HERE]

At this point in the training, students generally expressed much surprise at their improved performance, some shaking their heads as if to say, "How could I have been so dumb?", and others beaming as if to say, "Eureka!" Trainers seized this moment to state that the students now knew all that the trainer had to teach, and that the rest of the lessons would be "just practicing" what had been learned on "different stuff--some pictures, some reading."

2.3 Picture Task II (PT II)

The entire strategy routine, now shaped through the work with PT I, was consolidated through repetition with a new set of 20 stimulus cards. These were also chosen to form five mutually exclusive categories of four items each.

The student was encouraged to take the lead in replicating the routine with PT II. The trainers limited their role to eliciting recall of the steps of the routine (a) by asking, "What did you do before when [X was the case]?" and (b) by prompting the recall of specific techniques each student had already used with success. Trainers paid particular attention to:

(a) The group labels chosen by the student. Trainers asked what these labels were, if the student did not offer them first. If the labels were not workable for the category, the trainer asked, "Are you satisfied/happy with that group name?" and worked with the student to produce a better label.

(b) Keeping the group piles separated on the table once groups had been formed by the student. Trainers would remind students to do this, if necessary.

(c) Prompting students to use student-specific recall techniques for difficulties in recall. During recall, the trainer recorded the student's list on a sheet of paper, in groups enclosed in drawn rectangles. At the completion of recall, the trainer showed this chart to the student.

(d) Directing a student who had trouble in recalling a group or an item to "Move on to another group" or "Go on to the next group" and "Come back to this one later."

(e) Encouraging students, upon completion of recall, to turn over the piles of stimulus cards, one pile at a time, and thereby check their own recall. This was also done to cue "remembering" (recognition of) items not recalled.

Trainers finished up work with PT II by congratulating students on their performance, and indicating that now the student "owned" the routine, that it was part of his store of knowledge of "how to do things," and that "no one could take it from him."

2.4 The Sequence of Back-up Tasks

Students who did not meet the performance criterion for PT 1 were tracked into the back-up sequence. These were presented in order of decreasing difficulty. This sequence of increasingly easier tasks provided a continuum along which a beginning task could be located for these students. This procedure was intended to ensure that each student could demonstrate the basic grouping principle, even on very simple tasks, and then demonstrate competence on the more difficult grouping tasks. In essence, the training procedure sought to ensure that each student worked successfully on the criterion tasks, i.e., grouping or categorizing items, prior to applying the principle to the more difficult tasks.

2.4.1 Back-up Task 1 (BT-1)

BT-1 was the most difficult of the back-up tasks. BT-1 materials consisted of two groups of four stimulus cards each, which were presented to the student randomly shuffled. The format of the task was like that of PT 1. However, the student was presented with much less material. Teaching technique was the same as for PT 1. The criterion for successful performance was the recall of three out of four items for each of the two categories.

Students who met the criterion for BT-0 returned to the original sequence, attempting PT-1 once again. Students who did not meet the criterion for BT-0 continued on the back-up sequence, attempting BT-2. (See 2.4.2.)

2.4.2 Back-up Task 2

BT-2 consisted of two groups of four stimulus cards, each presented to the student in a blocked array. This array was arranged in a single horizontal line. If the student did not recognize the two groups in this line of cards, the trainers physically moved the cards so that a space was formed between the four items of the first group and the four items of the second group.

The performance criterion for BT-2 was the recall of three out of four items for each of the two groups. Students who met this criterion were returned to BT-1. Students who did not meet this criterion were tracked to BT-3.

2.4.3 Back-up Task 3 (BT-3)

BT-3 consisted of two groups of four stimulus cards each, presented to the student in a blocked array, and provided with group labels spoken by the trainers. For this task, the pre-grouped cards were once more arranged in a horizontal line. The trainers said:

> "Here are some [vegetables] and some [clothes]."
>
> "Can you show me which are [the vegetables]?"
>
> "Can you show me which are [the clothes]?"
>
> "Make a pile of all the [vegetable] cards."
>
> "Now make a pile of all the [clothes] cards."
>
> "Now let's see if you can remember which [vegetables] are in this small pile."

In this way, the tasks of grouping and memorizing were broken down into small "chunks" and the student was led through the "chunked" tasks by the trainer. Success at BT-3 was helped by the trainers' prompting, even at the stage of recall. Therefore, no student could fail to meet the performance criterion (three out of four items for each group) for this task. After completion of BT-3, the student was returned to BT-2, then BT-1 and PT-1.

2.5 Sequence of Prose Tasks

2.5.1 Bridge from the Picture Tasks

The trainer told the student that he had been "reading" the pictures on the cards in the previous lessons, and that both pictures and words are symbols we all agree on. The trainer continued:

> "You have been 'reading' pictures. And what you did with the pictures, you can do with words. So what we're going to do now is make more groups, but this time with things from reading."

Many students displayed a negative reaction when reading was mentioned. The trainers emphasized the transferability of the routine students had already successfully learned: "What you did with the pictures, you can do with the reading." Because: "Grouping is grouping. It's the same for both."

2.5.2 Preparatory Steps for Prose

The prose tasks required two preparatory steps prior to using the routine learned in the picture tasks. The first of these steps was, obviously, reading the passage. (Pr IA: See 2.5.2.2.) Students were asked to read aloud, "so that I [the trainer] can hear how you read." In this way, the trainer's awareness of the student's difficulties with reading was indirectly acknowledged. As each student completed the first reading, the trainer said: "You read much better than I thought you would." With this comment, the quality of the student's reading was dismissed by the trainer as a subject of conversation.

The second preparatory step for prose tasks was disembedding from the text items to be grouped. To initiate this step, the trainers asked the student a question which would separate items from text. For example, for a reading about some third graders' field trip to a farm, the trainer asked, "What did the third graders see at Drumlin Farm?" If a student had difficulty indicating discrete items, the trainer suggested that the student read the passage again, and pause to underline the words which would answer the trainer's question.

2.5.3 Prose Task I (PrI)

The first passage (PrI) presented selections in which items to be grouped had been blocked in the text. In addition, the text included a group label for each group. There were two passages in PrI, a short selection with three groups of four items each (PrIA), and a longer selection with five groups of four items each (PrIB). Both were required of all students.

After a student had read PrIA and had indicated that he could perform the disembedding step, the trainer cued the student to begin the grouping routine already learned.

> "Yes, those are things the children saw.
> Now put those things in order, just the way you

did with the cards."

Students were prompted to make grouped lists of the items, using rectangular borders to enclose each group.

Some students spontaneously grouped the items, either by associating items, or by beginning with the group names and filling in each group list. Students who had difficulty in grouping the disembedded words from the text were directed to make one long list containing all items. The trainer then questioned the student about the relation of the items, until the student produced usable groups (see 2.2.5).

> "Do any of these things go together?"
>
> ["Yes, the (car) and the (truck)."]
>
> "And what do you call (cars and trucks)?"

Some students needed to make a more obvious connection with the stimulus card task. For them, the trainer drew a series of boxes on a sheet of paper and wrote the disembedded words, one in each box. If necessary, these boxes could be cut out and handled exactly like the stimulus cards.

In PrI, group labels were included in the text. If the student did not spontaneously use these labels, the trainer directed him to re-read the text and look for the label.

> "This story tells you names for the groups.
> Read it again and see if you can find what the story calls these groups."

If a student produced a group name which differed from the label given in the text, the trainer asked for the text name and initiated a comparison. Whenever the student's name was useful, it was retained.

> "You have called this group the ________ group.
>
> What does the story say is the name of this group?
>
> Why did the story call the group _______? Why do you think?

which group name makes it easiest for you to remember this group?"

To review, the steps so far in the PrI task:

(a) Student reads the passage.

(b) Trainer asks a question which requires a list of items for the answer.

(c) Student disembeds the words which name these items.

(d) Group names (category labels) are found for each group, either from the student or from in the text.

(e) A written list of items is made in which the groups are labelled and separated by boundaries.

Learning of the group lists proceeded just as with the picture tasks. Once again, one group was memorized at a time. File cards or small pieces of paper were used to cover all group lists except the group the student was actively learning. Separate learning of the groups was monitored by the trainer; the trainer reminded the student to keep the group lists separated in the (metaphorical) "storage area" of the brain. Students more dependent on visual memory were taught an additional memonic technique. This technique consisted of imagining the group lists tacked up on the wall of the room, each group several feet from the others.

Again, the memorization was finished when the group lists were all "on their shelves, with the doors closed," and the student could produce a list of group names from memory as a separate unit.

The performance criterion for PrIA was the recall of three out of four items for two out of three groups. Failure to meet this criterion on the first attempt was followed by a second try at PrIA. This time the student was encouraged to find out what items were missed and to talk about how that happened. The emphasis for all students throughout was the maintenance of the group concept in recall. The failure to recall specific items was treated as a detail to be remedied by trial and error.

Prose Task IB was treated exactly like PrIA. All students who were proceeding at an average or above average pace completed PrIB. The performance criterion for PrIB was the recall of three out of four items for three out of five groups.

2.5.4 Prose Task II (PrIIA and PrIIB)

PrIIA and PrIIB presented the student with stories in which groups were blocked in the text; that is, the items of each group were found close together in the text. Unlike PrIA and PrIB, these readings did not supply the student with group labels. This difference was pointed out to the student by the trainers, who said:

> "Hmm. In the reading we did before, the one about X, the writer told us a name for the groups right in the story. In this story, you'll need to find some group names yourself--the way you did with the pictures. Sometimes you find group names in the reading, sometimes you don't."

The job of finding group labels was handled as in the picture tasks (see 2.2.5).

The performance criteria for these tasks were (a) for PrIIA: recall of three out of four items for two out of three groups; (b) for PrIIB: recall of three out of four items for three out of five groups.

As with PrI, all students completed PrIIA, and students who proceeded through the tasks at an average or above average rate completed PrIIB.

2.5.5 Prose Tasks III (PrIIIA and PrIIIB)

Prose task IIIA and IIIB presented the student with stories in which the items to be formed into groups were scattered throughout the stories in an unorganized fashion--a so-called "random" layout. Many students reacted to PrIII by commenting: "This doesn't make sense" or

"These are all mixed up."

In keeping with the teaching technique used throughout the training, trainers used questioning to elicit "spontaneous" formation of groups from random items. Group formation occurred either before or after a written list of items had been made by the student. Finding group names either accompanied the formation of groups, or took place as above (see 2.5.2.2) after the groups had been formed.

Learning then proceeded as above (see 2.2.5).

Criteria for successful performance for PrIIIA was the recall of three out of four items for three out of five groups. All students completed PrIIIA; students completed PrIIIB if time allowed.

2.5.6 Prose Task IV (PrIV)

PrIV was an optional task made available for those students who had progressed rapidly through the tasks, leaving extra time within the four session sequence. The story used in PrIV contained two mutually exclusive blocked groups of four items each. However, in order to answer the trainer's question "Is there enough information in your groups to let you tell the whole story?", these two groups had to be disassembled and a third group reconstituted from some of their members. This task represented a conceptual leap for students who had only been exposed to mutually exclusive grouping, and was intended for students who demonstrated a quick grasp of the concept of grouping in the previous sessions.

2.6 Making a List of Rules for Grouping

At the last session of the initial 4-week training sequence, the trainers told the students that they would like each student to have the rules for making groups written down, so that the student could keep them and use them whenever he wanted. Trainers emphasized that the students already had these steps well learned, that the student "owned" the rules, and that grouping was now "one of the things you know how to do, just way you know how to (ride a bike)."

The trainers questioned the student to elicit the steps of the well-practiced routine, shaping this list by filling in gaps and so forth, until it matched a master list of grouping rules (see list below). The language of this list varied with the student; simple language or more sophisticated language was used as was deemed appropriate to the student.

The completed list of rules was stapled into an oaktag folder and presented to the student with some ceremony.

2.6.1 "Rules for Making Groups"

1. Do the reading.
2. What question(s) do you want to answer?
   Ask yourself that question(s).
3. List the items.
4. Find the groups and name the groups.
   How many groups are there?
   How many things in each group?
5. Learn the groups.
   Remember to learn them separately.
6. Learn the list of group names.
7. Test yourself.
   Fill in any "lost" or "misplaced" items.

2.7 Conclusion of the Initial Training Sequence

Trainers used a checklist to ensure that each student left the initial training sequence with the following learning:

(a) Student has mastered a routine for the strategy.

  1) Student knows how to form groups under the various conditions.

  2) Student knows how to name groups with useful group names.

  3) Student has an active learning component in his strategy to store items learning, and realizes that learning improves with practice.

(b) Student is aware that knowing the process is the important learning, and that once he has learned the process it "belongs to him."

3.0 The Continued Training Sequence

3.1 The Long-term Condition Training

Students who were assigned to the long-term training condition worked with passages taken from the daily newspaper. These had been recast into the form of typewritten "stories," and their source was not known to the students. These students experienced the challenge of applying the grouping strategy to more difficult material. However, trainers did not draw the connection between grouping routine and everyday life for these students.

The nature of the long-term condition placed constraints on the teaching techniques used with these students. A question like "What would your teacher think is important in this reading?" (see 3.4) would have to be omitted for this group of students. When teaching long-term students, trainers relied on (a) the techniques used with the prose tasks of the first training sequence--underlining, listing, chart-making, etc.--and on (b) more extensive use of the construction of groups of attributes about the self as an analogue ("Facts about me"; see 3.4).

3.2 The Supportive Curriculum Condition

Students in the supportive curriculum condition were encouraged to generalize their group-making skills to written material encountered in everyday life, particularly in the classroom. In addition to articles reproduced from the daily paper, complete with photos and headlines, they analyzed reading material from their on-going classwork.

These circumstances afforded the trainers the chance to use whatever teaching opportunities presented themselves, with the aim of helping the students apply their strategy learning in the context of their everyday schoolwork. Various methods were used, including (a) discussing previous classroom lessons to decide what was "important" (see 3.4) in a passage, (b) making a chart of a student's family tree and relating this to a lesson on immigrants, and (c) examining automobiles in the school parking lot to clarify a newspaper story on auto inspection.

Trainers took advantage of whatever was helpful in clarifying a student's grasp of the material in any reading selection, and then they proceeded with the task of applying the strategy to the material.

3.3 Overview of Continued Training Sequence

The continued training sequence gave the students experience in transferring the strategy they had learned in the initial training sequence to the analysis of "real-life" printed material. Concepts taught in the continued training sequence (see 3.2) expanded on the original training by (a) building in flexibiity to the structure of the routine and (b) encouraging self-initiated, creative solutions to the task of grouping.

3.4 Concepts Taught in the Continued Training Sequence

Seven concepts were taught in this sequence. They were introduced and reinforced when teaching opportunities were presented by the materials on which the students worked. The order in which these concepts are discussed below was kept as a rough guide for the trainers. Because students' reading material varied in content throughout this sequence (see 3.1 and 3.2), this order was adjusted to achieve optimal synchronization between materials and teaching concepts.

The seven teaching concepts were:

(a) Categories, once made, can be disassembled and recategorized. This concept was learned by advanced students as the optional PrIV at the conclusion of the initial training sequence.

(b) Groups may be made up of any number of items. Groups from the same reading may be made up of various numbers of items. This important concept allowed students to escape from the "four items per group" expectation developed during initial training. The expectation of four items per group was helpful at the beginning of training, simplifying the students' group-formation. At this point in training, such an expectation acted as a hindrance to analysis of the texts.

(c) Events can be treated as items within a group. This concept greatly increased the scope of possible "items" for grouping. Groups in initial training were made up of concrete objects, with the exception of a few items which involved action upon concrete objects. For example, "She swept the floor and watered the plants." Students associated these actions to the object acted upon, and in initial training the different nature of the items was not commented upon by the trainer.

(d) "Facts about X" can be a group label.

(e) "Miscellaneous" can be a group label. These group labels provided additional organizing tools, and also introduced the notion that a category can be based on an idiosyncratic rationale. To introduce students to (d) "Facts about X," trainers had them make a list of "facts about themselves"; divided into groups or kept as one group these facts were listed under the group label: "Facts about [student's name]."

To explain the functional nature of (e) "Miscellaneous" as a group label, trainers used the metaphor of a drawer or box the student might have in his house, into which he puts all the things that are left over after he cleans up.

> "All the dishes are on the shelf, all the clothes are in the bureau drawers, all the trash is in the wastebasket, and you have a few safety pins, a roll of tape, a comb, some batteries, and a pair of dice left on the floor. You don't have any special place for these things, so you put them in a drawer of your bureau, and call it your "miscellaneous drawer."

(f) Nested categories, in which a larger group contains a smaller group or groups. The simple nested category introduced the students to the notion of hierarchical organization of grouped items. The chart which a student may have already made, entitled "Facts about [myself]" was sometimes used, either as is or in expanded form, to teach the "nested" concept. (See illustration.)

(g) Extracting "important" items from the reading versus extracting "everything" from the reading.

As the readings became more lengthy and complex, students came to realize that the amount of material in their groupings could become so enlarged as to be unmanageable. Trainers maintained the theory that any amount of material could be grouped and memorized given enough time, but asked aloud, "For what reason would anyone want to do this?" At this point, the focus of the routine became a functional one. Trainers directed students to first decide what the purpose of the reading was for them, before asking the appropriate question(s) of themselves and beginning their strategy routine.

The question of "what is important" information in a piece of reading was taken a step further with those students who could manage it. These students were encouraged to think about relative importance and importance from various perspectives, i.e., inference. Trainers asked them:

> "What do you think is more [or less] important to remember in this reading?"
>
> "What do you think the person who wrote this believes is important to remember in the story?"

> "What do you think your teacher would want you to remember in this reading?" (With SC condition; see 3.2.)

In deciding when to include the question of relative importance, the trainers used their judgment, as they had done before with the optional prose tasks. If relative importance was broached with a student and appeared to confuse him, the subject was dropped with the disclaimer, "Well, this is something you can think about later. Right now I'd like you to group the things you think are important in this story."

BEST COPY AVAILABLE